目　　录

第一章 $\mathbb{R}^n$ 中的凸集

我们假定，读者已经熟悉数学分析和线性代数的基本概念和基本知识，以及集合论中的常用符号和基本事实. 为了方便读者阅读，同时也为了统一符号，在 1.1 节和 1.2 节中我们扼要地叙述今后要用到的线性代数和数学分析中的一些重要事实. 本章讨论的主要内容包括：凸集和凸锥，凸集的相对内部，凸集分离定理，承托超平面，凸多面体，凸集的无界性和凸集的面. 凸集分离定理在凸分析中占有核心的作用，是凸分析的重要分析工具. 凸分析中的一些重要结论都与凸集分离定理密切相关.

1.1 仿射集和超平面

我们用 $\mathbb{R}$ 表示实数集，用 $\mathbb{R}^n$ 表示所有 n 实数组 $x = (\xi_1, \cdots, \xi_n)^\tau$ 构成的线性空间. x 叫做 $\mathbb{R}^n$ 的向量、点或元，这里 τ 表示转置. $\mathbb{R}^n$ 中向量的加法和数乘法规定如下：

$$(\xi_1, \cdots, \xi_n)^\tau + (\eta_1, \cdots, \eta_n)^\tau = (\xi_1 + \eta_1, \cdots, \xi_n + \eta_n)^\tau,$$
$$\lambda(\xi_1, \cdots, \xi_n)^\tau = (\lambda\xi_1, \cdots, \lambda\xi_n)^\tau,$$

其中 $\lambda \in \mathbb{R}$. $\mathbb{R}^n$ 中 m 个向量 $x_1, \cdots, x_m$ 叫做是线性无关的，是指对于实数 $\alpha_1, \cdots, \alpha_m$, $\alpha_1 x_1 + \cdots + \alpha_m x_m = 0 \Longrightarrow \alpha_i = 0, \forall i = 1, \cdots, m$. 集合 $M \subset \mathbb{R}^n$ 叫做子空间，是指它对于加法和数乘运算是封闭的，即

$$x, y \in M, \lambda \in \mathbb{R} \Longrightarrow x + y \in M, \lambda x \in M.$$

子空间 M 中线性无关向量的最大数目叫做该子空间的维数，记作 $\dim M$. 例如，$\mathbb{R}^n$ 的维数正好是 n，故 $\mathbb{R}^n$ 是一个 n 维空间. m 维子空间 M 中的 m 个线性无关向量 $x_1, \cdots, x_m$ 形成 M 的一个基，这时 M 中的任意向量 x 均可唯一地表示成

$$x = \alpha_1 x_1 + \cdots + \alpha_m x_m,$$

而相应的系数 $\alpha_1, \cdots, \alpha_m$ 叫做 x 在基 $x_1, \cdots, x_m$ 之下的坐标. 特别地，在 $\mathbb{R}^n$ 中，记 e_k 是第 k 个分量为 1 而其余分量为 0 的向量，则 $e_1, \cdots, e_n$ 是 $\mathbb{R}^n$ 的一个基，而 $x = (\xi_1, \cdots, \xi_n)^\tau$ 恰好是 x 在此基之下的坐标表现.

设 $M \subset \mathbb{R}^n$ 为任一集合. 包含 M 的最小子空间叫做由 M 张成的子空间，记作 $\operatorname{span} M$. 显然

$$\operatorname{span} M = \left\{ \sum_{k=1}^{m} \lambda_k x_k \;\middle|\; x_k \in M, \alpha_k \in \mathbb{R}, m \text{ 任意} \right\}.$$

对于 $\mathbb{R}^n$ 中的两个集合 M 和 N，以及实数 λ，我们定义 M 和 N 的和，以及 M 和 λ 的数乘如下:

$$M + N = \{x + y \mid x \in M, y \in N\},$$
$$\lambda M = \{\lambda x \mid x \in M\}.$$

容易验证上述运算满足如下分配律:

$$\lambda(M + N) = \lambda M + \lambda N; \quad (\lambda + \mu)M \subseteq \lambda M + \mu M.$$

给定 $\mathbb{R}^n$ 中两个不同的点 x 和 y，集合

$$\{(1 - \lambda)x + \lambda y \mid \lambda \in \mathbb{R}\}$$

叫做通过点 x 和 y 的直线，而集合

$$[x,y]=\{(1-\lambda)x+\lambda y \mid 0\le\lambda\le 1\}$$

叫做连接点 x 和 y 的闭线段. 类似地可以定义开线段 (x,y) 和半开半闭线段 $[x,y),(x,y]$ 等.

子集 $M\subset \mathbb{R}^n$ 称为仿射集或线性流形，是指通过 M 中任意两点的直线都在 M 中，即

$$\forall\, x,y\in M,\forall\,\lambda\in\mathbb{R}\Longrightarrow(1-\lambda)x+\lambda y\in M,$$

或者等价地，

$$(1-\lambda)M+\lambda M\subset M,\quad \forall\,\lambda\in\mathbb{R}.$$

空集 $\emptyset$ 和全空间 $\mathbb{R}^n$ 是仿射集的极端例子.

定理 1.1.1 为了集合 $M\subset\mathbb{R}^n$ 是一个子空间，必须且只须 M 是包含原点 0 的一个仿射集.

证明 每个子空间都含原点，并且显然是一个仿射集. 现在设 M 是含原点的仿射集. 对于任意 $x\in M$, $\lambda\in\mathbb{R}$, 我们有

$$\lambda x=(1-\lambda)0+\lambda x\in M,$$

即 M 对于数乘运算是封闭的. 今若 $x,y\in M$, 则

$$(1/2)(x+y)=x/2+y/2\in M.$$

因此依据上述数乘运算的封闭性，

$$x+y=2(x/2+y/2)\in M,$$

即 M 对于加法运算也封闭，从而 M 是一个子空间. ∎

对于集合 $M\subset\mathbb{R}^n$ 和 $a\in\mathbb{R}^n$, 称集合

$$M+a\triangleq\{x+a\mid x\in M\}$$

为 M 沿向量 a 的平移. 显然，仿射集的平移仍然是一个仿射集.

我们称仿射集 M 平行于仿射集 S,是指存在一向量 $a \in \mathbb{R}^n$, 使得 $M = S + a$. 显然这样定义的平行关系在 $\mathbb{R}^n$ 的所有仿射子集所构成的集族上是一种等价关系，换句话说，如果这种平行关系记作 $\|$, 则 (M, S, L 均为仿射集)

(1) $M \parallel M$(自反性);

(2) $M \parallel S \Longrightarrow S \parallel M$(对称性);

(3) $M \parallel S$, $S \parallel L \Longrightarrow M \parallel L$(传递性).

定理 1.1.2 $\mathbb{R}^n$ 中每个非空仿射集 M 都平行于一个 (唯一的) 子空间 L, L 表示成

$$L \triangleq M - M = \{x - y \mid x \in M, y \in M\}.$$

证明 首先证明 M 不可能平行于两个不同的子空间 L_1 和 L_2. 事实上，子空间 L_1 和 L_2 同时平行于 M 导致 L_1 和 L_2 彼此平行. 于是存在 $a \in \mathbb{R}^n$ 使 $L_2 = L_1 + a$. 由于 $0 \in L_2$, 故 $a \in L_1$. 因此 $L_1 \supset L_1 + a = L_2$. 类似地可证 $L_2 \supset L_1$, 从而 $L_1 = L_2$. 为了得到 L 的表达，只需注意对任意 $y \in M$, $M - y$ 是 M 的含原点 0 的一个平移，从而由定理 1.1.1 和刚才所证的结论，这个仿射集必定是与 M 平行的唯一的子空间 L. 由于 $y \in M$ 的任意性，实际上我们得到了 $L = M - M$. ∎

定理 1.1.2 表明，任意仿射集 M 是某个子空间 L 的一个平移，而且这个子空间 L 与 M 平行.

对于一个非空仿射集 M, 我们定义它的维数为与其相平行的这个子空间 L 的维数，即 $\dim M = \dim L$. 维数为 0,1 和 2 的仿射集分别叫做点，直线和平面. $\mathbb{R}^n$ 中的 $n-1$

维仿射集叫做超平面. 超平面在凸分析理论中起着很重要的作用. 为了进一步研究超平面的结构，我们先在 $\mathbb{R}^n$ 中引入内积 (或称标量积):

$$\langle x, y\rangle = \xi_1\eta_1 + \cdots + \xi_n\eta_n, \quad \forall\, x = (\xi_i), y = (\eta_i) \in \mathbb{R}^n.$$

关于内积的性质将在 1.2 节中详细讨论.

大家知道， $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性变换 A 是一个 $m \times n$ 矩阵，而 $\mathbb{R}^n$ 上的线性函数或线性泛函是 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的一个线性变换，并且 $\mathbb{R}^n$ 上的任意线性泛函 f 都能表示成 $f(x) = \langle x, b\rangle$, 其中 $b \in \mathbb{R}^n$ 由 f 唯一地确定. 通常我们把 f 等同于 $\mathbb{R}^n$ 中的一个向量.

我们称 $\mathbb{R}^n$ 中两个向量 x 和 y 彼此正交，记作 $x \perp y$, 是指 $\langle x, y\rangle = 0$. 给定一个集合 $M \subset \mathbb{R}^n$, 向量 $x \in \mathbb{R}^n$ 叫做与 M 正交，记作 $x \perp M$, 是指 $\langle x, y\rangle = 0, \forall\, y \in M$. $\mathbb{R}^n$ 中与 M 正交的所有向量的集合叫做 M 的正交补，记作 $M^\perp$. 显然 $M^\perp$ 是 $\mathbb{R}^n$ 的一个子空间. 如果 L 是 $\mathbb{R}^n$ 的一个子空间，则

$$\dim L + \dim L^\perp = n, \quad L = (L^\perp)^\perp.$$

如果 $b_1, \cdots, b_m$ 是子空间 L 的基，则

$$L^\perp = \{x \in \mathbb{R}^n \mid \langle x, b_k\rangle = 0,\ 1 \le k \le m\}.$$

特别 $\mathbb{R}^n$ 中 $n-1$ 维子空间 L 是某个一维子空间的正交补. 设 b 是该一维子空间的基，则

$$L = \{x \in \mathbb{R}^n \mid \langle x, b\rangle = 0\}.$$

根据定理 1.1.1, $\mathbb{R}^n$ 中任意超平面是 $n-1$ 维子空间的平移. 因此，对于 $\mathbb{R}^n$ 中的任意超平面 H, 存在两个向量

$a,b\in\mathbb{R}^n$, 使得

$$H=\{x\in\mathbb{R}^n\mid\langle x,b\rangle=0\}+a=\{y\in\mathbb{R}^n\mid\langle y,b\rangle=\beta\},$$

其中 $\beta=\langle a,b\rangle$. 这样我们证明了下述定理:

定理 1.1.3 $\mathbb{R}^n$ 中的每个超平面 H 都可以表示成

$$H=\{x\in\mathbb{R}^n\mid\langle x,b\rangle=\beta\},$$

其中 $b\in\mathbb{R}^n$, $\beta\in\mathbb{R}$, 并且 b 和 β 在不记及非零公共因子的意义下是唯一的.

在定理 1.1.3 中的向量 b 叫做超平面 H 的法向量. H 的别的法向量或者与 b 同方向, 或者与 b 反方向.

显然任意多个仿射集的交仍然是一个仿射集. 对于给定的集合 $S\subset\mathbb{R}^n$, 设 $x_1,\cdots,x_m\in S$, $\lambda_k\in\mathbb{R}$ 满足 $\lambda_1+\cdots+\lambda_m=1$, 则称向量 $x=\lambda_1x_1+\cdots+\lambda_mx_m$ 为诸向量 $x_1,\cdots,x_m$ 的一个仿射组合. 我们把包含 S 的最小的一个 (唯一的) 仿射集叫做 S 的仿射包, 记作 $\operatorname{aff}S$. 不难证明 (见习题 1.1.7), $\operatorname{aff}S$ 由 S 中向量的一切仿射组合构成, 即

$$\operatorname{aff}S=\left\{\sum_{i=1}^m\lambda_ix_i\,\middle|\,x_k\in S,\lambda_k\in\mathbb{R},\sum_{i=1}^m\lambda_i=1,\ m\ \text{任意}\right\}.$$

$\mathbb{R}^n$ 中 $m+1$ 个向量 $b_0,\cdots,b_m$ 叫做仿射无关的, 是指向量 $b_1-b_0,\cdots,b_m-b_0$ 线性无关; 而当 $b_1-b_0,\cdots,b_m-b_0$ 线性无关时就称 $m+1$ 个向量 $b_0,\cdots,b_m$ 是仿射相关的.

显然, 当 $m+1$ 个向量 $b_0,\cdots,b_m$ 仿射无关时,

$$\operatorname{aff}\{b_0,\cdots,b_m\}=L+b_0,$$

其中 L 为子空间,

$$\begin{aligned}L&=\operatorname{aff}\{0,b_1-b_0,\cdots,b_m-b_0\}\\&=\operatorname{span}\{b_1-b_0,\cdots,b_m-b_0\}.\end{aligned}$$

下面两个定理的证明是很容易的，其证明留作练习.

定理 1.1.4 设 $b_0,\cdots,b_m$ 是 $\mathbb{R}^n$ 中 $m+1$ 个向量，那么，为了 $b_0,\cdots,b_m$ 仿射相关，必须且只须存在不同时为零的 $m+1$ 个实数 $\lambda_0,\cdots,\lambda_m$，使得

$$\lambda_0+\cdots+\lambda_m=0;\quad \lambda_0 b_0+\cdots+\lambda_m b_m=0.$$

定理 1.1.5 给定 $\mathbb{R}^n$ 中 $m+1$ 个向量 $b_0,\cdots,b_m$，如果 $m\geq n+1$，则 $b_0,\cdots,b_m$ 必仿射相关.

设 $b_0,\cdots,b_m$ 是 $\mathbb{R}^n$ 中 $m+1$ 个仿射无关的向量. 点 $x\in\mathbb{R}^n$ 的关于 $b_0,\cdots,b_m$ 的仿射组合表示 $x=\lambda_0 b_0+\cdots+\lambda_m b_m$ 是唯一的. 事实上，如果 x 还能表示成仿射组合 $x=\mu_0 b_0+\cdots+\mu_m b_m$，则 $(\lambda_0-\mu_0)b_0+\cdots+(\lambda_m-\mu_m)b_m=0$. 但 $b_0,\cdots,b_m$ 是仿射无关的，故 $\lambda_k=\mu_k$, $\forall 0\leq k\leq m$. x 的这个唯一的仿射组合表示中的系数 λ_k 叫做 x 的仿射坐标.

由上述讨论可知，由 $m+1$ 个仿射无关的向量 $b_0,\cdots,b_m$ 所张成的仿射包 $\text{aff}\{b_0,\cdots,b_m\}$ 是一个 m 维的仿射集；反之，一个 m 维的仿射集均可表示成 $m+1$ 个点的仿射包.

称一个单值映射 $T:\mathbb{R}^n\to\mathbb{R}^m$ 为一个仿射变换，是指它满足

$$T((1-\lambda)x+\lambda y)=(1-\lambda)Tx+\lambda Ty,\ \forall\, x,y\in\mathbb{R}^n, \forall\,\lambda\in\mathbb{R}.$$

定理 1.1.6 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的仿射变换 T 可以表示成 $Tx=Ax+a$，其中 A 为 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性变换，而 a 为 $\mathbb{R}^m$ 的固定元.

证明 令 $a=T0$, $Ax=Tx-a$. 显然 A 仍是一个仿射变换，并且 $A0=0$. 于是，对于任意 $\lambda\in\mathbb{R}$, $x\in\mathbb{R}^n$,

$$A(\lambda x)=A(\lambda x+(1-\lambda)0)=\lambda Ax+(1-\lambda)A0=\lambda Ax,$$

即 A 是齐性的. 然后利用 A 的齐性, 对于 $x, y \in \mathbb{R}^n$,

$$A(x+y) = 2A(x/2 + y/2),$$

即 A 还是加法的. 从而 A 是 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的一个线性变换. 反之, 对于任意线性变换 $A: \mathbb{R}^n \to \mathbb{R}^m$, $Tx = Ax + a$ 显然是仿射的. ∎

定理 1.1.7 设 $b_0, \cdots, b_m$ 和 $a_0, \cdots, a_m$ 是 $\mathbb{R}^n$ 中的两组仿射无关的向量. 那么存在一个 $\mathbb{R}^n$ 到自身的一对一的仿射变换 T, 使得 $Tb_k = a_k$, $k = 0, 1, \cdots, m$. 如果 $m = n$, 则 T 还是唯一的.

证明 不妨对 $m = n$ 的情况证明之 (否则可以把原来的 $m+1$ 个仿射无关的向量扩大成 $n+1$ 个仿射无关向量). 由线性代数知道, 存在 $\mathbb{R}^n$ 的其自身的唯一的一对一的线性变换 A, 使得 $A(b_k - b_0) = a_k - a_0$, $k = 1, \cdots, n$. 于是 $Tx = Ax + (a_0 - Ab_0)$ 就是所要求的仿射变换. ∎

推论 1.1.8 设 M_1 和 M_2 是 $\mathbb{R}^n$ 中任意两个维数相同的仿射集, 那么存在一个 $\mathbb{R}^n$ 到其自身的一对一的仿射变换 T 使得 $TM_1 = M_2$.

习　题　1.1

1.1.1 设 H 是 $\mathbb{R}^n$ 中的超平面, b 是 H 的法向量. 试证 b 与位于 H 中的任意向量正交.

1.1.2 给定超平面 $H_1 = \{x \in \mathbb{R}^n \mid \langle x, b\rangle = \beta\}$ 和 $H_2 = a + H_1$, 其中 $a \in \mathbb{R}^n, \beta \in \mathbb{R}$. 试证: $H_2 = \{x \in \mathbb{R}^n \mid \langle x, b\rangle = \alpha\}$, 式中 $\alpha = \langle a, b\rangle + \beta$.

1.1.3 设 $H = \{x \in \mathbb{R}^2 \mid \langle x, b\rangle = 2\}$ 是 $\mathbb{R}^2$ 中的超平面, 其中 $b = (2, -3)^\tau$.

(1) 试画出超平面 H 和其法向量 b;

(2) 设 $x_0 = (-1, 1)^\tau$, 试画出超平面 $x_0 + H$ 的图并找出 $\beta \in \mathbb{R}$ 使得 $x_0 + H = \{x \in \mathbb{R}^2 \mid \langle x, b\rangle = \beta\}$.

1.1.4 试证 $\mathbb{R}^n$ 中每个 m 维仿射集 M 都可以表示成 $M = \{x \in \mathbb{R}^n \mid Bx = b\}$, 其中 $b \in \mathbb{R}^m$, B 是某个 $(n-m) \times n$ 矩阵.

1.1.5 试证 $\mathbb{R}^n$ 中每个仿射集都是有穷多个超平面的交.

1.1.6 设 L 是 $\mathbb{R}^n$ 的 $k(\leq n)$ 维子空间, 并设 $M_1 = x_1 + L$, $M_2 = x_2 + L$, 其中 $x_1, x_2 \in \mathbb{R}^n$. 试证：或者 $L_1 = L_2$, 或者 $L_1 \cap L_2 = \emptyset$.

1.1.7 设 $S \subset \mathbb{R}^n$, 试证:

$$\operatorname{aff} S = \left\{ \sum_{k=1}^{m} \lambda_k x_k \;\middle|\; x_k \in S, \lambda_k \in \mathbb{R}, \sum_{k=1}^{m} \lambda_k = 1 \right\}.$$

1.1.8 完成定理 1.1.4 的证明.

1.1.9 完成定理 1.1.5 的证明.

1.1.10 完成推论 1.1.8 的证明.

1.1.11 设 $M = \{(2,0)^\tau, (0,5)^\tau, (-1,1)^\tau\} \subset \mathbb{R}^2$. 试证:

(1) M 的三个向量是仿射无关的;

(2) 求出向量 $p_1 = (2,1)^\tau, p_2 = (1,1)^\tau, p_3 = (1,1/3)^\tau, p_4 = (1,0)^\tau$ 相对于 M 的仿射坐标.

1.1.12 假定 M_1 和 M_2 是 $\mathbb{R}^n$ 中的两个维数相同的仿射集, 如果 $M_1 \subset M_2$, 试证 $M_1 = M_2$.

1.1.13 在 $\mathbb{R}^4$ 中设 $p_1 = (1,-1,2,-1)^\tau$, $p_2 = (2,-1,2,0)^\tau$, $p_3 = (1,0,2,0)^\tau$, $p_4 = (1,0,3,1)^\tau$, 试证:

(1) 向量 p_1, p_2, p_3, p_4 仿射无关;

(2) 设 $A = \operatorname{aff}\{p_1, p_2, p_3, p_4\}$, $B = \{(\xi_1, \xi_2, \xi_3, \xi_4)^\tau \mid \xi_1 + \xi_2 + \xi_3 - \xi_4 = 3\}$, 则 $A = B$.

1.1.14 设 $x_1, \cdots, x_m$ 是 $\mathbb{R}^n$ 中的相互正交的非零向量, 试证它们是线性无关的.

1.2 $\mathbb{R}^n$ 中的拓扑结构

上一节我们定义了 $\mathbb{R}^n$ 中的内积 $\langle\cdot,\cdot\rangle$. 通常把定义了

内积的 $\langle\cdot,\cdot\rangle$ 的 n 维线性空间 $\mathbb{R}^n$ 叫做欧氏空间. 容易验证 $\mathbb{R}^n$ 中的内积 $\langle\cdot,\cdot\rangle$ 具有下列性质: 对于任意 $x, y \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, 我们有

(1) $\langle x, x\rangle$, 并且 $\langle x, x\rangle = 0 \Longleftrightarrow x = 0$;

(2) $\langle x, y\rangle = \langle y, x\rangle$;

(3) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$;

(4) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$.

我们可以通过内积来规定 $\mathbb{R}^n$ 中向量 x 的 "长度": $\|x\| = \langle x, x\rangle^{1/2}$, 叫做 x 的欧氏范数.

定理 1.2.1 $\mathbb{R}^n$ 中的范数 $\|\cdot\|$ 具有如下基本性质:

(1) $\|x\| \geq 0$, 并且 $\|x\| = 0 \Longleftrightarrow x = 0$;

(2) $\|\alpha x\| = |\alpha|\,\|x\|$;

(3) $\|x + y\| \leq \|x\| + \|y\|$;

(4) $|\langle x, y\rangle| \leq \|x\|\,\|y\|$,

这里 $x, y \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$.

证明 (1) 和 (2) 从定义直接得出. 为证 (4), 注意,

$$\|\alpha x + y\|^2 = \alpha^2\|x\|^2 + 2\alpha\langle x, y\rangle + \|y\|^2 \geq 0, \ \forall \alpha \in \mathbb{R}.$$

并且 $\|\alpha x + y\|^2 = \alpha^2\|x\|^2 + 2\alpha\langle x, y\rangle + \|y\|^2$ 作为 α 的二次函数的判别式

$$|\langle x, y\rangle|^2 - \|x\|^2\|y\|^2 \leq 0,$$

这正是 (4). 最后, (3) 是 (4) 的直接结果. ∎

对于 $x, y \in \mathbb{R}^n$, 我们定义 x 与 y 之间的距离为

$$d(x, y) = \|x - y\|.$$

定理 1.2.2 $\mathbb{R}^n$ 中的距离函数 $d(\cdot,\cdot)$ 具有如下性质: $\forall\, x, y \in \mathbb{R}^n, \lambda \in \mathbb{R}$,

(1) $d(x,y) \geq 0$, 并且 $d(x,y)=0 \Longleftrightarrow x=y$;

(2) $d(x,y)=d(y,x)$;

(3) $d(x+y,z) \leq d(x,z)+d(y,z)$;

(4) $d(\lambda x, \lambda y)=|\lambda|\, d(x,y)$;

(5) $d(x+z,y+z)=d(x,y)$.

证明 留作练习. ∎

对于任意点 $x \in \mathbb{R}^n$, 以 x 为中心、 ε 为半径的开球是指 $\mathbb{R}^n$ 中的集合

$$B(x,\varepsilon) \triangleq \{y \in \mathbb{R}^n \mid \|y-x\| < \varepsilon\}.$$

设集合 $M \subset \mathbb{R}^n$. 点 $x \in \mathbb{R}^n$ 叫做 M 的内点, 是指存在一正数 $\varepsilon > 0$ 使得 $B(x,\varepsilon) \subset M$; M 叫做开集是指 M 的每一点都是其内点; M 叫做闭集是指 M 的补集 $M^c = \mathbb{R}^n \backslash M$ 是开集.

容易验证, 开球、全空间 $\mathbb{R}^n$ 和空集 $\emptyset$ 是开集, 任意多个开集的并是开集, 有穷多个开集的交仍是开集。同样, 闭球 $\overline{B}(x,\varepsilon)=\{y \in \mathbb{R}^n \mid \|y-x\| \leq \varepsilon\}$, 全空间 $\mathbb{R}^n$ 和空集 $\emptyset$ 是闭集, 有限点集是闭集, 任意多个闭集的交以及有穷多个闭集的并仍然是闭集. 注意 $\mathbb{R}^n$ 中既不开又不闭的集合是有的, 例如 $\mathbb{R}$ 中的半开半闭区间即是这样的例子.

集合 $M \subset \mathbb{R}^n$ 的内部是指由 M 的所有内点组成的集合, 记作 $\operatorname{int} M$. 容易证明, $\operatorname{int} M$ 正好是所有包含 M 的开集的并.

点 $x \in \mathbb{R}^n$ 叫做集合 $M \subset \mathbb{R}^n$ 的附着点, 是指对于任意 $\varepsilon > 0$, $B(x,\varepsilon) \cap M \neq \emptyset$; x 叫做 M 的聚点或极限点, 是指 $B(x,\varepsilon) \cap (M \backslash \{x\}) \neq \emptyset$, $\forall \varepsilon > 0$.

集合 $M \subset \mathbb{R}^n$ 的闭包是指其所有附着点组成的集合, 记作 $\operatorname{cl} M$ (许多书中记作 $\overline{M}$). 容易证明 M 的闭包 $\operatorname{cl} M$

正好是所有包含 M 的闭集的交.

$\mathbb{R}^n$ 中的序列或点列 (x_k) 叫做以 $x \in \mathbb{R}^n$ 为极限或收敛于 x, 记作 $\lim_{k\to\infty} x_k = x$ 或 $x_k \to x\ (k \to \infty)$, 是指它满足

$$\lim_{k\to\infty} \|x_k - x\| = 0.$$

集合 $M \subset \mathbb{R}^n$ 的边界 $\mathrm{bd}\, M$ 是指 M 的闭包与其补集 M^c 的闭包的交集, 即

$$\mathrm{bd}\, M = \mathrm{cl}\, M \cap \mathrm{cl}\, M^c.$$

(有的书中把集合 M 的边界记作 ∂M, 但本书中 ∂F 将表示函数 F 的次微分, 见第四章) 从定义容易看出, 为了点 $x \in \mathrm{bd}\, M$, 必须且只须 $\forall\, \delta > 0$, 开球 $B(x, \delta)$ 与 M 和 M^c 都相交, 即

$$B(x,\delta) \cap M \neq \emptyset,\ B(x,\delta) \cap M^c \neq \emptyset,\ \ \forall\, \delta > 0.$$

集合 $M \subset \mathbb{R}^n$ 称为有界的, 是指它包含在 $\mathbb{R}^n$ 的某个球中. M 叫做相对紧的, 是指 M 的任意无限子集含有极限点; 如果这样的极限点一定在 M 中, 则称 M 是紧集. 大家知道, 根据分析中著名的 Weierstrass 定理, M 为相对紧集的充分必要条件是 M 为有界集, 而 M 为紧集的充分必要条件是 M 为有界闭集.

刻划 $\mathbb{R}^n$ 中集合紧性的另一个重要定理是所谓有穷覆盖定理. 对于集合 $M \subset \mathbb{R}^n$, $\mathbb{R}^n$ 中的集合族 $\{O_i \mid i \in I\}$(I 为某个指标集) 叫做 M 的一个覆盖, 是指它满足 $M \subset \cup\{O_i \mid i \in I\}$; 此外, 若每个集合 O_i 还是开的, 则此覆盖称为开覆盖.

定理 1.2.3 为了集合 $M \subset \mathbb{R}^n$ 是紧集, 必须且只须对于 M 的任意一族开覆盖 $\{O_i \mid i \in I\}$, 必能从中找出一个有限开覆盖覆盖 M, 即存在有限个标号 $i_1, \cdots, i_m \in I$ 使得

$$M \subset \cap\{O_{i_k} \mid k=1,\cdots,m\}.$$

这个定理是数学分析中的基本结果，读者可查阅有关的教科书.

$I\!R^n$ 中的一个集合族 $\mathcal{F}=\{F_\lambda \mid \lambda\in\Lambda\}$ 叫做具有有穷交性质，是指 $\mathcal{F}$ 的任意有穷子族具有非空交. 从定理 1.2.3 可得如下定理：

定理 1.2.4 集合 $M\subset I\!R^n$ 为紧集的充分必要条件是包含在 M 中的具有有穷交性质的任意闭集族均有非空交.

证明 先证必要性. 设 M 为紧集，而 M 中的一个闭集族 $\mathcal{F}=\{F_\lambda \mid \lambda\in\Lambda\}$ 具有有穷交性质. 我们要证明 $\bigcap\limits_{\lambda\in\Lambda} F_\lambda\neq\emptyset$. 事实上，若不然，则 $\cap\{F_\lambda \mid \lambda\in\Lambda\}=\emptyset$, 从而

$$\cup\{F_\lambda^c \mid \lambda\in\Lambda\}=I\!R^n\supset M,$$

其中 $F_\lambda^c=I\!R^n\backslash F_\lambda$ 为 F_λ 在 $I\!R^n$ 中的补集. 这样，由紧集的有穷覆盖定理 1.2.3, 存在有穷个标号 $\lambda_k\in\Lambda, k=1,\cdots,m$ 使得

$$\cup\{F_{\lambda_k}^c \mid k=1,\cdots,m\}\supset M,$$

或者等价地

$$\cap\{F_{\lambda_k} \mid k=1,\cdots,m\}\subset M^c.$$

由此可见 $\cap\{F_{\lambda_k} \mid \lambda\in\Lambda\}=\emptyset$. 这与 $\mathcal{F}$ 的有穷交性质相矛盾. 现在证明充分性. 假定包含在 M 中具有有穷交性质的每个闭集族均具有非空交. 设 $\{O_\lambda \mid \lambda\in\Lambda\}$ 为 M 的一个开覆盖, $\cup\{O_\lambda \mid \lambda\in\Lambda\}\supset M$. 于是 $O_\lambda^c\cap M$ 为闭集, 并且

$$\cap\{O_\lambda^c\cap M \mid \lambda\in\Lambda\}=\emptyset.$$

因此， M 中闭集族 $\{O_\lambda^c\cap M \mid \lambda\in\Lambda\}$ 必定没有有穷交性质，从而存在有穷个标号 $\lambda_k\in\Lambda, k=1,\cdots,m$ 使得

$$M\cap\left(\cap\{O_{\lambda_k}^c \mid k=1,\cdots,m\}\right)=\emptyset,$$

即

$$M^c \cup \left(\cup \{O_{\lambda_k} \mid k = 1, \cdots, m\} \right) = \mathbb{R}^n.$$

因此 $M \subset \cup\{O_{\lambda_k}^c \mid k = 1, \cdots, m\}$. ∎

推论 1.2.5 设 $\mathcal{F}$ 是 $\mathbb{R}^n$ 中一族紧集. 如果 $\mathcal{F}$ 具有有穷交性质，则 $\mathcal{F}$ 必具有非空交.

函数或映射的连续性是分析中的重要概念. 我们说函数 $f : \mathbb{R}^n \to \mathbb{R}^m$ 在点 $x_0 \in \mathbb{R}^n$ 是连续的，是指对于任意 $\varepsilon > 0$, 存在 $\delta > 0$, 使得 $f(B(x_0, \delta) \subset B(f(x_0), \varepsilon)$. 这里为简单起见，不同维数空间中的球都不加区别地记作 $B(\cdot, \cdot)$.

下列两个定理的证明留作练习.

定理 1.2.6 设 $f : \mathbb{R}^n \to \mathbb{R}^m$ 是连续函数， M 是 $\mathbb{R}^n$ 的紧子集，则 $f(M) = \{f(x) \mid x \in M\}$ 是 $\mathbb{R}^m$ 中的紧子集.

定理 1.2.7 设 $f : \mathbb{R}^n \to \mathbb{R}$ 是连续函数， M 是 $\mathbb{R}^n$ 的紧子集，则 $f(M) = \{f(x) \mid x \in M\}$ 是 R 中的有界闭集，并且 f 在 M 上达到其上确界和下确界，即存在 $x_1, x_2 \in M$, 使得

$$f(x_1) = \inf\{f(x) \mid x \in M\},$$

$$f(x_2) = \sup\{f(x) \mid x \in M\}.$$

函数 $f : \mathbb{R}^n \to \mathbb{R}$ 叫做在 $x_0 \in \mathbb{R}^n$ 处下半连续，是指对于任意序列 $(x_k) \subset \mathbb{R}^n$, $x_k \to x_0 (k \to \infty)$, 有

$$\varliminf_{k \to \infty} f(x_k) \geq f(x_0);$$

f 称为在 x_0 处上半连续，是指 $-f$ 在 x_0 处下半连续. 容易看出， f 在 x_0 处的下半连续性等价于：对于任意点列 $(x_k) \subset \mathbb{R}^n$ 和数列 $(\mu_k) \subset \mathbb{R}$, 如果满足

$$\lim_{k \to \infty} \mu_k = \mu, \ \mu_k \geq f(x_k), \ \forall, k \geq 1,$$

则必有 $\mu \geq f(x_0)$.

定义在子集上的函数的连续性可类似地定义.

最后我们简单地提一下点集的稠密性. 设 M 和 S 是 $I\!R^n$ 的两个子集. 我们说集合 S 在 M 中稠密，简称稠，是指 $M \subset \mathrm{cl}\, S$, 换句话说, M 中的元可以用 S 中的元任意逼近，确切地说，对于任意元 $z \in M$, 必存在序列 $(x_k) \subset S$, 使得 $x_k \to z(k \to \infty)$.

例如, $I\!R$ 中的全体有理数集在 $I\!R$ 中稠；同样地, $I\!R^n$ 中所有有理坐标点的集合在 $I\!R^n$ 中也是稠密的.

习 题 1.2

1.2.1 完成定理 1.2.2 的证明.

1.2.2 试证如下几个集合都是开集:

(1) 开球 $B(x,\varepsilon)$;

(2) $I\!R^n$;

(3) $\emptyset$;

(4) 任意多个开集的并集;

(5) 有穷多个开集的交集.

1.2.3 试证如下几个集合都是闭集:

(1) 闭球 $\overline{B}(x,\varepsilon)$;

(2) $I\!R^n$;

(3) $\emptyset$;

(4) 任意多个闭集的交集;

(5) 有穷多个闭集的并集.

1.2.4 试证: M 的内部 $\mathrm{int}\, M$ 恰好是所有包含在 M 中的开集的并集.

1.2.5 试证: M 的闭包 $\mathrm{cl}\, M$ 恰好是所有包含 M 的闭集的交集.

1.2.6 试证: 为了 x 是 M 的极限点，必须且只须存在一个点列 $(x_k) \subset M \backslash \{x\}$ 使得 $x_k \to x(k \to \infty)$.

1.2.7 试证：$\mathbb{R}^n$ 中的点列收敛等价于按每个坐标分量收敛. 换句话说，如果 $x_k = (\xi_{k1}, \cdots, \xi_{kn})^\tau$, 则 $x_k \to x(k \to \infty)$ 等价于对每个 $i, 1 \le i \le n, \xi_{ki} \to \xi_i(k \to \infty)$, 其中 $x = (\xi_1, \cdots, \xi_n)^\tau$.

1.2.8 试证：函数 $f : \mathbb{R}^n \to \mathbb{R}^m$ 在 $x_0 \in \mathbb{R}^n$ 处连续的充分必要条件是，对于 $\mathbb{R}^n$ 中任意收敛于 x_0 的点列 (x_k), 都有 $f(x_k) \to f(x_0)$ $(k \to \infty)$.

1.2.9 对于下列命题，或者证它成立，或者用反例指出它不成立：

(1) 若 A 是开集，B 是任意集合，则 $A + B$ 也是开集.

(2) 若 A 和 B 都是闭集，则 $A + B$ 也是闭集.

1.2.10 设 A 和 B 是 $\mathbb{R}^n$ 中的非空集合，我们定义 A 到 B 的距离为 $d(A, B) \triangleq \inf\{d(x, y) \mid x \in A, y \in B\}$. 如果 A 是单点集，$A = \{x\}$, 则 $d(x, B) \triangleq d(\{x\}, B)$ 叫做点 x 到 B 的距离.

(1) 试证：若 B 是闭集，$x \notin B$, 则 $d(x, B) > 0$; 并且 $d(x, B)$ 作为 $x \in \mathbb{R}^n$ 的函数在 $\mathbb{R}^n$ 上是连续的.

(2) 举一个例子说明对于不相交的两个闭子集 A 和 B, 有可能 $d(A, B) = 0$.

(3) 试证：若 A 是闭集，B 是紧集，并且 $A \cap B = \emptyset$, 则 $d(A, B) > 0$.

1.2.11 给定集合 $M \subset \mathbb{R}^n$, 试证 $x \in \operatorname{bd} M$ 当且仅当 $\forall \delta > 0$, $B(x, \delta) \cap M \ne \emptyset$, $B(x, \delta) \cap M^c \ne \emptyset$.

1.2.12 完成定理 1.2.6 和 1.2.7 的证明.

1.2.13 利用定理 1.2.3, 证明定理 1.2.4 成立.

1.2.14 完成推论 1.2.5 的证明.

1.2.15 设 S 是 $\mathbb{R}^n$ 中的开子集. 试证：对于任意给定的闭子集 $M \subset S$, 存在 $\mathbb{R}^n$ 的一个开子集 T 使得 $M \subset T \subset \operatorname{cl} T \subset S$.

1.2.16 设 M 是 $\mathbb{R}^n$ 中的有界闭子集. 试证：对于任意 $\varepsilon > 0$, 必存在 M 中有穷个点 $x_1, \cdots, x_m$ 使得 $M \subset \cup\{B(x_j, \varepsilon) \mid j = 1, \cdots, m\}$ (这里的 $\{x_1, \cdots, x_m\}$ 叫做 M 的有穷 ε- 网); 并且如果 S 是 M 的稠子集，则上述有穷个点可以在 S 中取.

1.2.17 设 $S \subset \mathbb{R}^n$ 为有界集. 试证：$\operatorname{cl}(\operatorname{conv} S) = \operatorname{conv}(\operatorname{cl} S)$.

特别当 S 是闭有界集时， $\operatorname{conv} S$ 也是闭有界的.

1.3 凸集和凸锥

凸集和凸函数是凸分析研究的主要对象. 本节介绍凸集的基本概念. 一般 n 维空间凸集概念是平面上凸集概念的自然推广.

$\mathbb{R}^n$ 中的一个子集 M 叫做凸的，是指它包含连接其中任意两点的线段. 显然，凸集 M 的特征可以表示成：

$$(1-\lambda)x+\lambda y\in M,\quad \forall x,y\in M, \forall 0\le\lambda\le 1,$$

或者等价地，

$$(1-\lambda)M+\lambda M\subset M,\quad \forall 0\le\lambda\le 1.$$

一个凸集可以是有界的，也可以是无界的 (见图 1.3.1). 下面是凸集的一些简单的例子.

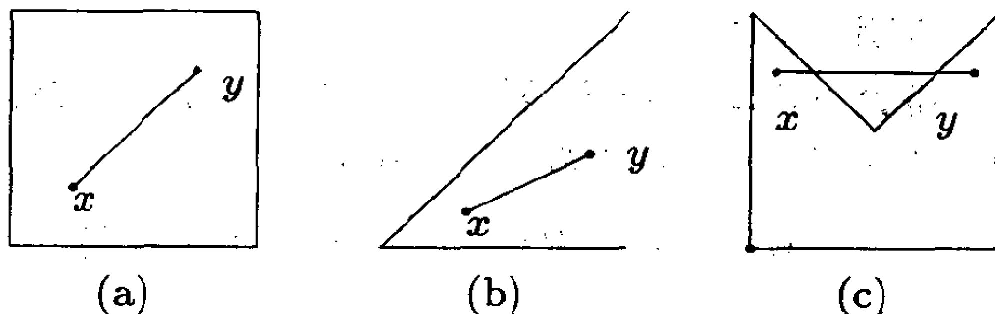


图 1.3.1 (a) 有界凸集； (b) 无界凸集； (c) 非凸集

例 1.3.1 平面上的线段、三角形和圆盘都是凸集.

例 1.3.2 $\mathbb{R}^n$ 中的仿射集是凸集.

例 1.3.3 $\mathbb{R}^n$ 中的半空间. 对于任意非零向量 $b\in\mathbb{R}^n$

和任意实数 $\beta \in \mathbb{R}$, 集合

$$\{x \in \mathbb{R}^n \mid \langle x, b\rangle \le \beta\},\ \{x \in \mathbb{R}^n \mid \langle x, b\rangle \ge \beta\}$$

叫做闭半空间；而集合

$$\{x \in \mathbb{R}^n \mid \langle x, b\rangle < \beta\},\ \{x \in \mathbb{R}^n \mid \langle x, b\rangle > \beta\}$$

叫做开半空间. 这四个集合显然是非空凸集，并且开半空间是开集，而闭半空间则是闭集. 注意，如果对某个实数 $\lambda \ne 0$, 用 λb 和 $\lambda\beta$ 分别代替 b 和 β, 则得到同样的四个半空间. 因此这些半空间只依赖于超平面 $H = \{x \in \mathbb{R}^n \mid \langle x, b\rangle = \beta\}$.

例 1.3.4 $\mathbb{R}^n$ 中的正象限

$$\{x = (\xi_1, \cdots, \xi_n)^\tau \in \mathbb{R}^n \mid \xi_k > 0, \forall\, 1 \le k \le n\}$$

和非负象限

$$\{x = (\xi_1, \cdots, \xi_n)^\tau \in \mathbb{R}^n \mid \xi_k \le 0, \forall\, 1 \le k \le n\}$$

都是凸集，并且前者是开集，而后者是闭集.

定理 1.3.1 任意多个凸集的交集仍然是凸集.

证明 留作练习. ∎

推论 1.3.2 设 I 是任一指标集. 对于每个指标 $i \in I$, 给定 $b_i \in \mathbb{R}^n$ 和 $\beta_i \in \mathbb{R}$, 那么集合

$$M = \{x \in \mathbb{R}^n \mid \langle x, b_i\rangle \le \beta_i, \forall\, i \in I\}$$

是凸集.

证明 记 $M_i = \{x \in \mathbb{R}^n \mid \langle x, b_i\rangle \le \beta_i\}$, 则 M_i 或者是一个半空间，或者是全空间 $\mathbb{R}^n$, 或者是空集 $\emptyset$, 从而总是凸集. 但 $M = \bigcap_{i\in I} M_i$, 所以由定理 1.3.1 知 M 是凸集. ∎

如果上述不等号分别用 $\ge$, $>$, $<$ 或 $=$ 代替，显然推论 1.3.2 的结论仍然是成立的.

对于 $\mathbb{R}^n$ 中的向量 $x_1, \cdots, x_m$, 以及实数 $\lambda_i \geq 0, \lambda_1 + \cdots + \lambda_m = 1$, 向量 $\lambda_1 x_1 + \cdots + \lambda_m x_m$ 叫做 $x_1, \cdots, x_m$ 的凸组合. 今后为了简单起见，当谈到向量 $x_1, \cdots, x_m$ 的凸组合 $\lambda_1 x_1 + \cdots + \lambda_m x_m$ 时，往往省略掉不言而喻的条件 $\lambda_i \geq 0, \lambda_1 + \cdots + \lambda_m = 1$.

定理 1.3.3 为了子集 $M \subset \mathbb{R}^n$ 是凸集，必须且只须 M 包含其中向量的一切凸组合.

证明 实际上，根据定义，集合 M 是凸的当且仅当 M 包含其中任意两个元的凸组合. 我们必须由此推出 M 包含其中任意多个元的凸组合. 为此，我们采用数学归纳法. 假定 M 包含其中任意 $m \geq 2$ 个元的凸组合. 现在任意取 M 中 $m+1$ 个元 $x_1, \cdots, x_{m+1}$ 的凸组合

$$x = \lambda_1 x_1 + \cdots + \lambda_{m+1} x_{m+1}.$$

不妨设 $0 < \lambda_{m+1} < 1$ (否则的话， x 成为 m 元的凸组合，依据归纳假设， $x \in M$)，于是若记 $\lambda = \lambda_1 + \cdots + \lambda_m$，则 $0 < \lambda < 1$，而且 $\lambda_{m+1} = 1 - \lambda$. 令 $\lambda'_k = \lambda_k / \lambda$，则 $\lambda'_k \geq 0$, $\lambda'_1 + \cdots + \lambda'_m = 1$. 因此依据归纳假设， $y = \lambda'_1 x_1 + \cdots + \lambda'_m x_m \in M$. 另一方面， x 可以表示成 $x = \lambda y + (1-\lambda) x_{m+1}$，于是由 M 的凸性知 $x \in M$. ▮

给定集合 $M \subset \mathbb{R}^n$，我们把包含 M 的最小的凸集叫做 M 的凸包，记作 $\operatorname{conv} M$. 从定义可以看出， M 的凸包 $\operatorname{conv} M$ 正好是所有包含 M 的凸集的交集.

定理 1.3.4 给定任意集合 $M \subset \mathbb{R}^n$，凸包 $\operatorname{conv} M$ 是由 M 中元的所有凸组合组成的集合.

证明 用 S 表示 M 中元的所有凸组合组成的集合. 依据定理 1.3.3，显然有 $S \subset M$. 任取 $x, y \in S$，则存在 $x_i, y_j \in M$, $1 \leq i \leq m$, $1 \leq j \leq k$，使得 x 和 y 表示成凸

组合 $x = \lambda_1 x_1 + \cdots + \lambda_m x_m$, $y = \mu_1 y_1 + \cdots + \mu_k y_k$. 于是，对于 $0 < \lambda < 1$,

$$(1-\lambda)x + \lambda y = \sum_{i=1}^{m}(1-\lambda)\lambda_i x_i + \sum_{j=1}^{k}\lambda\mu_j y_j$$

是 M 中元的一个凸组合，从而 $(1-\lambda)x + \lambda y \in S$. 这就是说 S 是一个凸集，并且 $M \subset S$, 这样我们证明了 $S = \operatorname{conv} M$. ∎

推论 1.3.5 $\mathbb{R}^n$ 中有限子集 $\{b_0, \cdots, b_m\}$ 的凸包由形如 $\lambda_0 b_0 + \cdots + \lambda_m b_m$ 的凸组合全体组成.

$\mathbb{R}^n$ 中有限点集的凸包叫做凸多面体，也简称多面体. 如果 $b_0, \cdots, b_m$ 仿射无关，即 $\dim \operatorname{aff}\{b_0, \cdots, b_m\} = m$, 则 $\operatorname{conv}\{b_0, \cdots, b_m\}$ 叫做 m 维单纯形，而点 $b_0, \cdots, b_m$ 叫做这个单纯形的顶点. 当 $m = 0, 1, 2, 3$ 时，相应的单纯形分别叫做点、直线段、三角形和四面体.

下一个定理在有穷维空间中凸集理论研究中是一个基本的结果，是由 Caratheodory 在 1907 年证明的.

定理 1.3.6 设 M 是 $\mathbb{R}^n$ 中的一个非空集，那么 M 的凸包中的点可以表示成 M 中至多 $n+1$ 个点的凸组合.

证明 任取一点 $x \in \operatorname{conv} M$, 依据定理 1.3.4, x 可以表示成 M 中元的凸组合，即 $x = \lambda_1 x_1 + \cdots + \lambda_m x_m$. 我们的目的是证明 x 可以表示成 M 中 $m \leq n+1$ 个点的凸组合.

现在若 $m > n+1$, 依据定理 1.1.5, $x_1, \cdots, x_m$ 必仿射相关，从而存在不全为零的实数 $\alpha_1, \cdots, \alpha_m$, 使得

$$\alpha_1 + \cdots + \alpha_m = 0, \quad \alpha_1 x_1 + \cdots + \alpha_m x_m = 0.$$

我们不妨假定 $\alpha_m > 0$ (否则把标号重新排序), 并且对

于 $\alpha_k > 0$ 的标号 k 满足 $\lambda_m/\alpha_m \le \lambda_k/\alpha_k$. 对于 $1 \le k \le m$, 令 $\beta_k = \lambda_k - (\lambda_m/\alpha_m)\alpha_k$. 于是 $\beta_m = 0, \beta_k \ge 0$, $1 \le k \le m-1$, 并且

$$\sum_{k=1}^{m} \beta_k = \sum_{k=1}^{m} \lambda_k - \frac{\lambda_m}{\alpha_m} \sum_{k=1}^{m} \alpha_k = 1.$$

另一方面, 我们有

$$\begin{aligned}\sum_{k=1}^{m-1} \beta_k x_k = \sum_{k=1}^{m} \beta_k x_k &= \sum_{k=1}^{m} \left(\lambda_k - \frac{\lambda_m}{\alpha_m}\alpha_k\right) x_k \\ &= \sum_{k=1}^{m} \lambda_k x_k - \frac{\lambda_m}{\alpha_m} \sum_{k=1}^{m} \alpha_k x_k = x.\end{aligned}$$

这样我们把 x 表示成了 M 中 $m-1$ 个元的凸组合. 这个过程可以重复进行下去, 直至 x 表示成 M 中至多 $n+1$ 个元的凸组合. ∎

依据定理 1.3.4, 单纯形 $S = \operatorname{conv}\{b_0, \cdots, b_m\}$ 中任意点都可以表示成其顶点的凸组合. 其实这种表示还是唯一的. 事实上, 如果

$$x = \sum_{k=0}^{m} \alpha_k b_k = \sum_{k=0}^{m} \beta_k b_k,$$

则 $\sum\limits_{k=0}^{m} \lambda_k b_k = 0$, 其中 $\lambda_k = \alpha_k - \beta_k$, $0 \le k \le m$. 注意,

$$\sum_{k=0}^{m} \lambda_k = \sum_{k=0}^{m} \alpha_k - \sum_{k=0}^{m} \beta_k = 1 - 1 = 0,$$

于是 $\lambda_1(b_1 - b_0) + \cdots + \lambda_m(b_m - b_0) = 0$; 但是 $\dim S = m$, 所以 $b_1 - b_0, \cdots, b_m - b_0$ 线性无关. 由此可见 $\lambda_1 = \cdots =$

$\lambda_m = 0$, 即 $\alpha_k = \beta_k, 0 \le k \le m$. 这样我们证明了如下定理:

定理 1.3.7 $\mathbb{R}^n$ 中任意一个 m 维单纯形 $\text{conv}\{b_0, \cdots, b_m\}$ 中每一点表示成其顶点的凸组合的形式是唯一的.

对于 $\mathbb{R}^n$ 中 m 维单纯形 $\text{conv}\{b_0, \cdots, b_m\}$, 设其中的点 x 表示成其顶点的凸组合的形式是

$$x = \alpha_0 b_0 + \cdots + \alpha_m b_m,$$

则数 $\alpha_0, \cdots, \alpha_m$ 叫做 x 的重心坐标, 而点 $b = \frac{1}{m}(b_0 + \cdots + b_m)$ 叫做这个单纯形的形心.

$\mathbb{R}^n$ 中凸集 M 的维数是指 M 的仿射包 $\text{aff}\, M$ 的维数. 这样, 一根直线段的维数是 1, 一个圆盘和一个三角形的维数是 2, 而不管它们所在的空间的维数是多少.

定理 1.3.8 $\mathbb{R}^n$ 中凸集 M 的维数是包含在 M 中的各单纯形的最大维数.

证明 M 的任意子集的凸包包含在 M 中. 因此, 如果 M 中 $m+1$ 个元 $b_0, \cdots, b_m$ 仿射无关,则单纯形 $\text{conv}\{b_0, \cdots, b_m\} \subset M$. 于是假定 m 是使得 M 中 $m+1$ 个元仿射无关的最大数目, 并设 $b_0, \cdots, b_m$ 为 M 中 $m+1$ 个仿射无关元. 设 S 是其仿射包, 即 $S = \text{aff}\{b_0, \cdots, b_m\}$. 于是 $\dim S = m$, 并且 $S \subset \text{aff}\, M$. 我们有 $M \subset S$, 因为如果有元 $b \in M \backslash S$, 则 $m+2$ 元 $b_0, b_1 \cdots, b_m, b$ 将仿射无关, 这与 m 的最大性相矛盾. 但 $\text{aff}\, M$ 是包含 M 的最小的仿射集, 故 $S = \text{aff}\, M$. 这样我们证明了 $\dim M = \dim S = m$.∎

从上述定理可见, 当一个凸集 M 的维数是 m 时, 必定有一个 m 维单纯形包含在 M 中. 这一事实在下一节证明凸集的相对内部非空时是有用的.

今后当一个凸集 $M \subset \mathbb{R}^n$ 的内部非空时, 即 $\text{int}\, M \neq \emptyset$

时，往往称 M 是一个凸体.

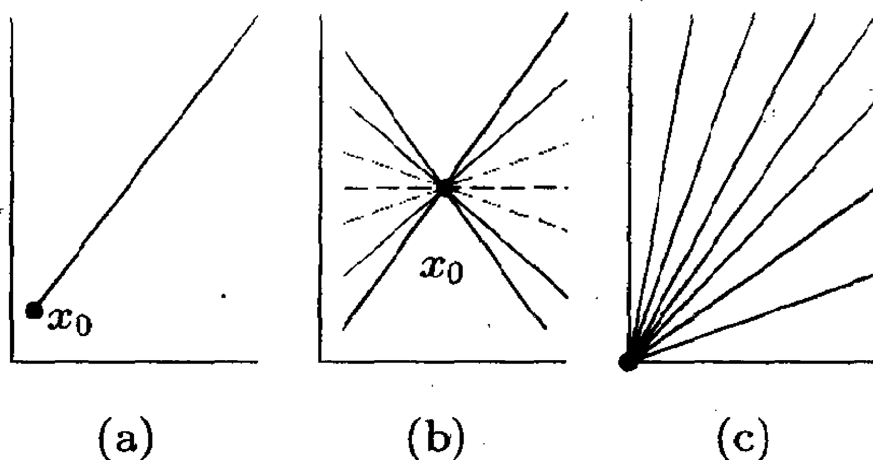


图 1.3.2　(a) 半直线，(b) 锥，(c) 凸锥

在这一节的最后，我们给出 $\mathbb{R}^n$ 中一类特殊的凸集，即所谓凸锥的定义及其基本性质. 给定 $\mathbb{R}^n$ 中两个不同点 x_0 和 x_1, 我们称集合

$$R(x_0;x_1)=\{x_0+\lambda(x_1-x_0)\mid\lambda>0\}$$

为从 x_0 出发的通过点 x_1 的半直线或射线. 如果 $\mathbb{R}^n$ 中的集合 K 满足

$$R(x_0;x)\subset K,\quad \forall\, x\in K,$$

则称 K 为 $\mathbb{R}^n$ 中以 x_0 为顶点的锥. 以 0 为顶点的锥简称为锥. 显然 K 为锥的充分必要条件是

$$\lambda x\in K,\quad \forall\, x\in K, \forall\,\lambda>0.$$

如果一个锥同时又是一个凸集，则称其为凸锥 (见图 1.3.2).

定理 1.3.9　集合 $M\subset\mathbb{R}^n$ 为凸锥的充分必要条件是

$$\lambda x+\mu y\in K,\quad \forall\,\lambda,\mu>0. \tag{1.3.1}$$

证明　首先设 K 为凸锥. 任取 $x,y\in K$, $\lambda,\mu>0$. 从 K 的凸性可知，

$$\frac{\lambda}{\lambda+\mu}x+\frac{\mu}{\lambda+\mu}y\in K,$$

而从 K 的锥性又可知

$$\lambda x+\mu y=(\lambda+\mu)\Big(\frac{\lambda}{\lambda+\mu}x+\frac{\mu}{\lambda+\mu}y\Big)\in K,$$

即式 (1.3.1) 成立. 反之，设式 (1.3.1) 成立. 于是特别地在式 (1.3.1) 中取 λ,μ 还满足 $\lambda+\mu=1$, 可知 K 为凸集. 而若取 $x=y\in K$, 则又可知 K 为锥. 从而 K 是凸锥. ∎

从凸锥定义直接可以得到如下定理：

定理 1.3.10 $\mathbb{R}^n$ 中任意一族凸锥的交仍是凸锥.

$\mathbb{R}^n$ 中的锥 K 叫做是真的，是指 $x\neq 0, x\in K \Longrightarrow -x\notin K$. 全空间 $\mathbb{R}^n$, $\mathbb{R}^n$ 中的正象限

$$\{x=(\xi_1,\cdots,\xi_n)^{\tau}\mid \xi_i>0, i=1,2,\cdots,n\}$$

和非负象限

$$\{x=(\xi_1,\cdots,\xi_n)^{\tau}\mid \xi_i\geq 0, i=1,2,\cdots,n\}$$

都是凸锥的例子，而且后面两个还是真凸锥.

设 $b_i\in\mathbb{R}^n$, $i\in I$ (I 为任意指标集). 那么显然

$$K=\{x\in\mathbb{R}^n\mid \langle x,b_i\rangle\leq 0,\forall i\in I\}$$

是凸锥. 这样，齐次不等式组的解集是一个凸锥.

定理 1.3.11 设 $M\subset\mathbb{R}$ 是任一非空子集，K 为 M 的所有正线性组合 (即在线性组合 $\lambda_1x_1+\cdots+\lambda_mx_m$ 中诸系数 λ_i 都是正的) 构成的集合. 那么 K 是包含 M 的最小的凸锥.

证明 K 是凸锥并且 $K \supset M$ 是显然的. 又根据定理 1.3.9, 任何一个包含 M 的凸锥都应该包含 M 的一切正线性组合, 从而 K 是包含 M 的最小的凸锥. ▮

推论 1.3.12 设 $M \subset \mathbb{R}$ 是凸子集, 并设

$$K = \{\lambda x \mid \lambda > 0, x \in M\},$$

那么 K 是含 M 的最小的凸锥.

证明 只需注意 M 中元的每个线性组合必定是 M 中元的凸组合的正倍数. ▮

定理 1.3.11 中的锥加上原点后得到的凸锥叫做由 M 生成的凸锥, 记作 $K = \operatorname{cone} M$, 即

$$\operatorname{cone} M = \left\{ \sum_{i=1}^{m} \lambda_i x_i \mid x_i \in M, \lambda_i \geq 0, 1 \leq i \leq m, m\text{任意} \right\}.$$

习 题 1.3

1.3.1 完成定理 1.3.1 的证明.

1.3.2 设 A 和 B 是 $\mathbb{R}^n$ 中的凸集, 试证 $A + B$ 也是凸集.

1.3.3 设 M 是 $\mathbb{R}^n$ 中的凸集, 并且 $\lambda_1 \geq 0, \lambda_2 \geq 0$, 求证:

$$(\lambda_1 + \lambda_2)M = \lambda_1 M + \lambda_2 M.$$

1.3.4 找出 $\mathbb{R}^2$ 中满足如下条件的点 $(x, y)^\tau$ 的集合的凸包:

(1) $y = x^2$;

(2) $y = x^2, x \geq 0$;

(3) $y = \cos x$;

(4) $y = 1/x, x \geq 1/2$.

1.3.5 试证有界集的凸包仍是有界的.

1.3.6 设 A, B 是 $\mathbb{R}^n$ 中的非空集合. 试证下列命题成立:

(1) $A \subset B \Longrightarrow A \subset \operatorname{conv} B$.

(2) $A \subset B$ 并且 B 凸 $\Longrightarrow \operatorname{conv} A \subset B$.

(3) $A \subset B \Longrightarrow \operatorname{conv} A \subset \operatorname{conv} B$.

(4) $(\operatorname{conv} A) \cup (\operatorname{conv} B) \subset \operatorname{conv}(A \cup B)$.

(5) $\operatorname{conv}(A \cap B) \subset \operatorname{conv} A \cap \operatorname{conv} B$.

(6) $\operatorname{conv}(\operatorname{conv} A) = \operatorname{conv} A$.

1.3.7 完成定理 1.3.10 的证明.

1.3.8 设 M 是 $\mathbb{R}^n$ 中的凸集, 试证: $\operatorname{bd}(\operatorname{cl} M) = \operatorname{bd} M$, 并举例说明这里 M 的凸性是必要的.

1.3.9 设 A, B 为 $\mathbb{R}^n$ 中的凸集, 试证:

$$\operatorname{conv}(A \cup B) = \{\lambda A + (1 - \lambda)B \mid 0 \le \lambda \le 1\}.$$

1.3.10 设 $T : \mathbb{R}^n \to \mathbb{R}^m$ 是线性变换, 如果 $A \subset \mathbb{R}^n$ 是凸子集, 试证: $T(A)$ 是 $\mathbb{R}^m$ 中的凸子集.

1.3.11 试证: $\mathbb{R}^n$ 中的任意一族仿射集的交仍是仿射集, 但任意一族仿射集的并却未必是仿射集或凸集.

1.3.12 设 M 是 $\mathbb{R}^n$ 中的凸集. 如果 $\operatorname{aff} M = \mathbb{R}^n$, 试证:

$$\operatorname{int} M \neq \emptyset.$$

1.3.13 设 M_1 和 M_2 是 $\mathbb{R}^n$ 中的仿射集. 试证: $M_1 \cup M_2$ 是凸集, 当且仅当 $M_1 \subset M_2$, 或者 $M_2 \subset M_1$.

1.3.14 在 $\mathbb{R}^2$ 中设 $p_1 = (0,1)^\tau, p_2 = (1,3)^\tau, p_3 = (4,3)^\tau$, $p_4 = (4,0)^\tau$, 并且 $x = (7/4, 5/4)^\tau$. 试证:

$$x = \frac{1}{2}p_1 + \frac{1}{4}p_2 + \frac{1}{6}p_3 + \frac{1}{12}p_4,$$

并应用 Caratheodory 定理证明中的程序把 x 表示成 p_1, p_2 和 p_3 的凸组合.

1.3.15 试证 Caratheodory 定理的如下形式: 设 M 是 $\mathbb{R}^n$ 中的非空子集, 设 $y \in M$, $x \in \operatorname{conv} M$, 那么存在点 $x_1, \cdots, x_k \in M$ $(k \le n)$ 使得 $x \in \operatorname{conv}\{y, x_1, \cdots, x_k\}$.

1.3.16 设 M 是 $\mathbb{R}^n$ 中非负象限 $\mathbb{R}^n_+$ 中的闭子集, 并设 e_k 是第 k 个分量为 1 其余分量为 0 的向量. 假定对于任意 $x \in M$, 存在正

实数 $\lambda_1, \cdots, \lambda_n$ 使得 $x + \lambda_k e_k \in M, \forall 1 \le k \le n$. 试证: $\operatorname{conv} M$ 是闭子集.

1.4 凸集的相对内部

$\mathbb{R}^3$ 中的线段和三角形作为 $\mathbb{R}^3$ 中的集合其内部都是空集. 但显然它们都有着自然的内部, 这就是相对内部.

我们把 $\mathbb{R}^n$ 中的集合 M 看成其仿射包 $\operatorname{aff} M$ 中的子集, 然后在 $\operatorname{aff} M$ 中规定 M 的内点和内部, 就得到 M 的相对内点和相对内部. 确切地说, 点 x 叫做 M 的相对内点, 是指存在 $\varepsilon > 0$ 使得

$$B(x, \varepsilon) \cap \operatorname{aff} M \subset M.$$

M 的相对内点全体叫做 M 的相对内部, 记作 $\operatorname{ri} M$. 于是

$$\operatorname{ri} M = \{x \in M \mid \exists \varepsilon > 0 \text{ 使得 } B(x, \varepsilon) \cap \operatorname{aff} M \subset M\}.$$

差集 $\operatorname{cl} M \backslash \operatorname{ri} M$ 叫做 M 的相对边界, 记作 $\operatorname{bd_r} M$. 集合 M 叫做相对开的, 是指 $M = \operatorname{ri} M$. 当 $\operatorname{aff} M = \mathbb{R}^n$ 时, 集合 M 的相对内部与其普通内部相同, 即 $\operatorname{int} M = \operatorname{ri} M$.

例如, 设 x, y 为 $\mathbb{R}^n$ 中不同的两个点, $[x, y]$ 为连接 x 和 y 的闭线段, 则

$$\operatorname{ri}[x, y] = \{\lambda x + (1 - \lambda) y \mid 0 < \lambda < 1\} = (x, y).$$

定理 1.4.1 设 M 是 $\mathbb{R}^n$ 中的凸集, 则任意连接 M 的闭包 $\operatorname{cl} M$ 中点和 M 的相对内点的线段的相对内点均是 M 的相对内点, 换句话说, 对于 $x_0 \in \operatorname{cl} M$, $x_1 \in \operatorname{ri} M$, 我们有 $x(\lambda) \triangleq (1 - \lambda) x_0 + \lambda x_1 \in \operatorname{ri} M, \forall 0 < \lambda < 1$.

证明 从假设 $x_1 \in \operatorname{ri} M$ 可知存在 $\varepsilon > 0$ 使得

$$B(x_1, \varepsilon) \cap \operatorname{aff} M \subset M.$$

为了证明当 $0 < \lambda < 1$ 时 $x(\lambda) \in \operatorname{ri} M$, 我们要证存在 $\delta > 0$ 使得 $B(x(\lambda), \delta) \cap \operatorname{aff} M \subset M$. 由于 M 是凸集, 为此我们只要证明对于任意 $z \in B(x\lambda), \delta) \cap \operatorname{aff} M$, 存在 $u, v \in M$ 使得 $z = \lambda v + (1-\lambda)u$. 首先注意, 如果 $z = \lambda v + (1-\lambda)u$, 则

$$\begin{aligned} \|v - x_1\| &= \left\|\frac{z}{\lambda} - \frac{1-\lambda}{\lambda} u - x_1\right\| \\ &\leq \frac{1}{\lambda}\|z - x(\lambda)\| + \frac{1-\lambda}{\lambda}\|x_0 - u\|. \end{aligned}$$

现在由于 $x_0 \in \operatorname{cl} M$, 我们可以取一点 $u \in M$ 使得 $\|u - x_0\| < \frac{\lambda}{1-\lambda}\frac{\varepsilon}{2}$. 再取 $v = \frac{1}{\lambda}z - \frac{1-\lambda}{\lambda}u$, 则 $z = \lambda v + (1-\lambda)u$. 这样只要取 $\delta = \frac{\lambda\varepsilon}{2}$, 就有 $\|v - x_1\| < \varepsilon$. 再考虑到 $\frac{1}{\lambda} - \frac{1-\lambda}{\lambda} = 1$, 可知 $v \in \operatorname{aff} M$, 从而 $v \in M$. ∎

定理 1.4.2 设 M 是 $\mathbb{R}^n$ 中的凸集, 那么 $\operatorname{cl} M$ 与 $\operatorname{ri} M$ 都是 $\mathbb{R}^n$ 中的凸集, 并且 $\operatorname{cl} M$ 与 M 有相同的仿射包, 即 $\operatorname{aff}(\operatorname{cl} M) = \operatorname{aff} M$, 从而 $\operatorname{cl} M$ 与 M 的维数也相同.

证明 $\operatorname{cl} M$ 的凸性是显然的;而在定理 1.4.1 中取 $x_0, x_1 \in \operatorname{ri} M$, 即得 $\operatorname{ri} M$. 最后, 注意到 $\operatorname{cl} M \subset \operatorname{aff} M$, 立即推出 $\operatorname{aff} M = \operatorname{aff}(\operatorname{cl} M)$. ∎

推论 1.4.3 设 $S \triangleq \operatorname{conv}\{b_0, b_1, \cdots, b_m\}$ 是 $\mathbb{R}^n$ 中的 m 维单纯形. 那么 S 的形心 $b = (b_0 + \cdots + b_m)/(m+1)$ 是 S 的相对内点, 从而 $\operatorname{ri} S \neq \emptyset$.

证明 由于 $b_0, \cdots, b_m$ 仿射无关, $\operatorname{aff} S$ 中每一点都可以表示成 $b_0, \cdots, b_m$ 的唯一的仿射组合. 对于 $x = \alpha_0 b_0 + \cdots + \alpha_m b_m \in \operatorname{aff} S$, 这里 α_k 为 x 的仿射坐标, 定义映射 $f : \operatorname{aff} S \to \mathbb{R}^{n+1}$:

$$f(x) = (\alpha_0, \cdots, \alpha_m)^\tau.$$

特别对于 S 的形心 b, 有

$$f(b) = \left(\frac{1}{n+1}, \cdots, \frac{1}{n+1}\right)^\tau.$$

注意对于仿射组合 $x = \alpha_0 b_0 + \cdots + \alpha_m b_m = b_0 + \alpha_1(b_1 - b_0) + \cdots + \alpha_m(b_m - b_0)$, 由于 $b_1 - b_0, \cdots, b_m - b_0$ 线性无关，可知

$$\sum_{k=1}^{m} \alpha_k \langle b_k - b_0, b_j - b_0 \rangle = \langle x - b_0, b_j - b_0 \rangle, \ 1 \le j \le m$$

有唯一解 (α_k), 并且 α_k 对于 x 是连续的，从而 f 是 $\operatorname{aff} S$ 上的连续函数. 但 f 在 b 处的值 $f(b)$ 的分量都是正的，因此存在一个开球 $B(b, \varepsilon)$ 使得 $B(b, \varepsilon) \cap \operatorname{aff} S$ 中的每一点 x 是 $b_0, \cdots, b_m$ 的凸组合，从而 $B(b, \varepsilon) \cap \operatorname{aff} S \subset \operatorname{conv} S$, 即 b 是 S 的相对内点. ∎

定理 1.4.4 设 M 是 $\mathbb{R}^n$ 中的一个 k $(k \ge 0)$ 维凸集. 那么 M 的相对内部 $\operatorname{ri} M$ 是非空的.

证明 由于 M 是 k 维凸集，故存在 $k+1$ 个仿射无关的向量 $b_0, \cdots, b_k \in M$ 使得 $\operatorname{aff} M = \operatorname{aff} S$, 其中 $S = \operatorname{conv}\{b_0, \cdots, b_k\}$ 是 k 维单纯形. 根据引理 1.4.3, $\operatorname{ri} S \ne \emptyset$. 但显然 $\operatorname{ri} S \subset \operatorname{ri} M$, 从而 $\operatorname{ri} M \ne \emptyset$. ∎

定理 1.4.5 设 M 是 $\mathbb{R}^n$ 中的任意凸集，则我们有

$$\operatorname{ri}(\operatorname{cl} M) = \operatorname{ri} M, \quad \operatorname{cl}(\operatorname{ri} M) = \operatorname{cl} M.$$

证明 由于 $\operatorname{ri} M \subset M$, 显然有 $\operatorname{cl}(\operatorname{ri} M) \subset \operatorname{cl} M$. 现在设 $y \in \operatorname{cl} M$. 依据定理 1.4.4, $\operatorname{ri} M \ne \emptyset$ (除非 $M = \emptyset$), 从而可以取一点 $x \in \operatorname{ri} M$. 从定理 1.4.1 可知 $(1-\lambda)x + \lambda y \in$

$\operatorname{ri} M, \forall 0 \leq \lambda < 1$. 让 $\lambda \to 1$ 即得 $y \in \operatorname{cl}(\operatorname{ri} M)$. 这样我们证明了 $\operatorname{cl}(\operatorname{ri} M) = \operatorname{cl} M$.

注意到 $\operatorname{cl} M \supset M$ 并且 $\operatorname{aff}(\operatorname{cl} M) = \operatorname{aff} M$, 可知 $\operatorname{ri}(\operatorname{cl} M) \supset \operatorname{ri} M$. 现在设 $z \in \operatorname{ri}(\operatorname{cl} M)$, 我们要证 $z \in \operatorname{ri} M$, 即存在 $\varepsilon > 0$ 使 $B(z, \varepsilon) \cap \operatorname{aff} M \subset M$. 为此任取一点 $x \in \operatorname{ri} M$ (不妨认为 $x \neq z$, 否则已经有 $z \in \operatorname{ri} M$), 并考虑通过 x 和 z 的直线 $(1-\mu)x + \mu z, \mu \in \mathbb{R}$. 当 $\mu > 1$ 且 $\mu - 1$ 充分小时, 位于该直线上的点

$$y(\mu) = (1-\mu)x + \mu z = z - (\mu - 1)(x - z)$$

仍属于 $\operatorname{ri}(\operatorname{cl} M)$, 因此 $y(\mu) \in \operatorname{cl} M$. 对于这样的 $y(\mu)$, z 可以表示成 $z = (1-\lambda)x + \lambda y(\mu)$, 式中 $\lambda = 1/\mu$. 于是根据定理 1.4.1, $z \in \operatorname{ri} M$. ∎

推论 1.4.6 设 M_1 和 M_2 是 $\mathbb{R}^n$ 中的两个凸集, 那么

$$\operatorname{cl} M_1 = \operatorname{cl} M_2 \Longleftrightarrow \operatorname{ri} M_1 = \operatorname{ri} M_2 \Longleftrightarrow \operatorname{ri} M_1 \subset M_2 \subset \operatorname{cl} M_1.$$

证明 留作练习. ∎

推论 1.4.7 设 M 是 $\mathbb{R}^n$ 中的非空凸集, O 是 $\mathbb{R}^n$ 的任一开集. 如果 $O \cap \operatorname{cl} M \neq \emptyset$, 则 $O \cap \operatorname{ri} M \neq \emptyset$.

证明 留作练习. ∎

推论 1.4.8 设 M_2 是 $\mathbb{R}^n$ 中的非空凸集, 而 M_1 是 M_2 的相对边界的凸子集. 那么 $\dim M_1 < \dim M_2$.

证明 如果 M_1 的维数与 M_2 的维数相同, 则 M_1 应该有相对于 $\operatorname{aff} M_2$ 的内点. 这样的点必然也是 M_2 的相对内点, 但根据假设, $\operatorname{ri} M_2 \cap M_1 = \emptyset$, 导致矛盾. ∎

定理 1.4.9 设 M 是 $\mathbb{R}^n$ 中的非空子集, 那么,

1) 如果 M 是有界集, 则 $\operatorname{conv} M$ 也是有界集;

2) 如果 M 是相对开集, 则 $\operatorname{conv} M$ 也是相对开集.

证明 (1) 的证明留作练习. 今证 (2). 设 M 是相对开集. 从 $M \subset \operatorname{conv} M$ 且 $\dim M = \dim(\operatorname{conv} M)$ 可知 $M = \operatorname{ri} M \subset \operatorname{ri}(\operatorname{conv} M)$. 另一方面, 从定理 1.4.2 可知 $\operatorname{ri}(\operatorname{conv} M)$ 是凸集. 因此 $\operatorname{conv} M \subset \operatorname{ri}(\operatorname{conv} M)$, 即 $\operatorname{conv} M$ 是相对开集. ∎

一般说来, 闭集的凸包未必是闭的. 例如, 在 $\mathbb{R}^2$ 中考虑集合

$$M = \{(\xi, \eta)^\tau \in \mathbb{R}^2 \mid \xi^2\eta^2 = 1, \eta > 0\}.$$

不难看出 M 是 $\mathbb{R}^2$ 中的闭集, 但其凸包 $\operatorname{conv} M = \{(\xi, \eta)^\tau \mid \eta > 0\}$ 却不是闭的 (见图 1.4.1). 然而如果我们要求 M 既是闭的又是有界的, 则 M 的凸包也是闭有界的.

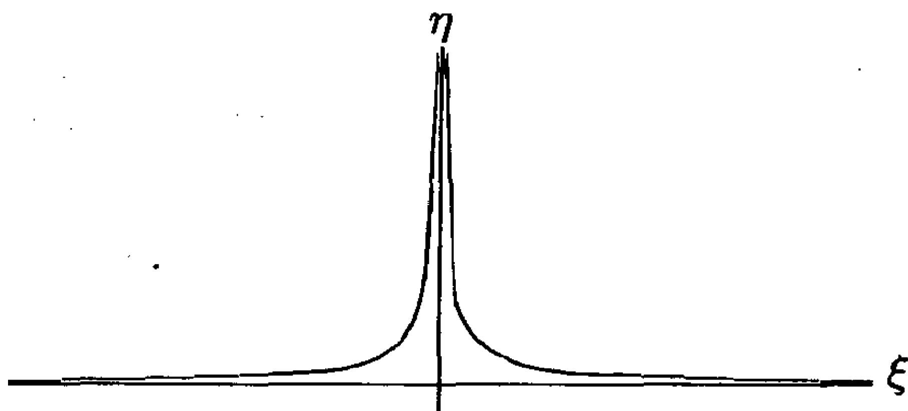


图 1.4.1

定理 1.4.10 设 M 是 $\mathbb{R}^n$ 中的紧子集, 那么 $\operatorname{conv} M$ 也是紧子集.

证明 我们定义 $\mathbb{R}^{n+1}$ 中的集合 S 如下:

$$S = \{(\alpha_1, \cdots, \alpha_{n+1})^\tau \mid \alpha_1 + \cdots + \alpha_{n+1} = 1, \alpha_i \geq 0, \}.$$

显然 S 是 $\mathbb{R}^{n+1}$ 中的有界闭集, 从而是紧集. 定义 $\mathbb{R}^{n+1} \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n = \mathbb{R}^{(n+1)^2}$ 到 $\mathbb{R}^n$ 的映射 f 如下:

$$f(\alpha_1,\cdots,\alpha_{n+1};x_1,\cdots,x_{n+1})=\sum_{k=1}^{n+1}\alpha_k x_k.$$

容易看出 $f:\mathbb{R}^{(n+1)^2}\to\mathbb{R}^n$ 是连续映射. 从 Caratheodory 定理 1.3.6 可知, f 把 $S\times\overbrace{M\times\cdots\times M}^{n+1\text{ 次}}$ 映到 $\operatorname{conv}M$ 之上. 但 $S\times\overbrace{M\times\cdots\times M}^{n+1\text{ 次}}$ 是 $\mathbb{R}^{(n+1)^2}$ 中的有界闭集, 从而 $\operatorname{conv}M=f\Big(S\times\overbrace{M\times\cdots\times M}^{n+1\text{ 次}}\Big)$ 是 $\mathbb{R}^n$ 中的有界闭集. ∎

下一个定理刻划了凸集的相对内点的特征, 在凸分析的一些理论证明中是十分有用的.

定理 1.4.11 设 M 是 $\mathbb{R}^n$ 中的一个非空凸集, 那么 $z\in\operatorname{ri}M$ 充分必要条件是: $\forall x\in M$ 存在 $\mu>1$ 使得 $(1-\mu)x+\mu z\in M$.

证明 定理中的条件是说, M 中以其相对内点 z 为一端点的直线段都可以延伸至 z 之外而不出 M. 必要性是显然的. 今证充分性:

设 z 满足定理条件. 首先从引理 1.4.3 知 $\operatorname{ri}M\neq\emptyset$, 从而存在点 $x\in M$. 由假设, 存在 $\mu>1$ 使得 $y\triangleq(1-\mu)x+\mu z\in M$. 于是 $z=(1-\lambda)x+\lambda y$, 其中 $0<\lambda=\mu^{-1}<1$. 因此依据定理 1.4.1, $z\in\operatorname{ri}M$. ∎

推论 1.4.12 设 M 是 $\mathbb{R}^n$ 中的非空凸集, 那么 $z\in\operatorname{int}M$ 当且仅当对于每一 $y\in\mathbb{R}^n$, 存在某个 $\varepsilon>0$ 使得 $z+\varepsilon y\in M$.

定理 1.4.13 设 $\{M_j\mid j\in J\}$ 是 $\mathbb{R}^n$ 中的一族凸集 (J 为一指标集). 假定 $\cap\{\operatorname{ri}M_j\mid j\in J\}\neq\emptyset$, 则

$$\mathrm{cl}\Big(\cap\{M_j \mid j\in J\}\Big)=\cap\{\mathrm{cl}\, M_j \mid j\in J\}. \tag{1.4.1}$$

此外，如果指标集 J 还是有限的，则还有

$$\mathrm{ri}\Big(\cap\{M_j \mid j\in J\}\Big)=\cap\{\mathrm{ri}\, M_j \mid j\in J\}. \tag{1.4.2}$$

证明 任取一点 $x\in\cap\{\mathrm{ri}\, M_j \mid j\in J\}$. 于是对于任意 $y\in\cap\{\mathrm{cl}\, M_j \mid j\in J\}$, 由定理 1.4.1 可知,

$$(1-\lambda)x+\lambda y\in\mathrm{ri}\, M_j,\quad \forall\,\lambda\in[0,1),\forall j\in J.$$

注意到 $y=\lim\limits_{\lambda\to 1-0}[(1-\lambda)x+\lambda y]$, 有

$$\begin{aligned}\cap\{\mathrm{cl}\, M_j \mid j\in J\}&\subset\mathrm{cl}\Big(\cap\{\mathrm{ri}\, M_j \mid j\in J\}\Big)\\&\subset\mathrm{cl}\Big(\cap\{M_j \mid j\in J\}\Big)\subset\cap\{\mathrm{cl}\, M_j \mid j\in J\},\end{aligned}$$

即式 (1.4.1) 成立. 同时上述一系列包含关系也证明了 $\cap\{M_j \mid j\in J\}$ 和 $\cap\{\mathrm{ri}\, M_j \mid j\in J\}$ 有相同的闭包. 这样，由定理 1.4.5 可知此两集必有相同的相对内部. 因此,

$$\mathrm{ri}\,(\cap\{M_j \mid j\in J\})\subset\cap\{\mathrm{ri}\, M_j \mid j\in J\}.$$

现在假定 J 为有限集. 任取一点 $z\in\cap\{\mathrm{ri}\, M_j \mid j\in J\}$, 依据定理 1.4.8, $\cap\{M_j \mid j\in J\}$ 中以 z 为一端点的每一条直线段都可以稍稍延伸至 z 之外，而同时又都保持在每个集合 M_j 之中. 但 J 是有限集，故这些延伸了的线段之交仍然是原来那个线段的延伸，并且落在 $\cap\{M_j \mid j\in J\}$ 之中. 由此依据定理 1.4.8 的判据， $z\in\mathrm{ri}\,(\cap\{M_j \mid j\in J\})$. 于是我们证明了

$$\cap\{\mathrm{ri}\, M_j \mid j\in J\}\subset\mathrm{ri}\,(\cap\{M_j \mid j\in J\}).$$ ∎

定理 1.4.14 设 M 是 $\mathbb{R}^n$ 中的凸集，并设 $A:\mathbb{R}^n\to\mathbb{R}^m$ 为一个线性变换. 那么,

$$\mathrm{ri}\,(AM) = A(\mathrm{ri}\,M), \quad \mathrm{cl}\,(AM) \supset A(\mathrm{cl}\,M).$$

证明 关于闭包运算的包含关系是 A 的连续性的直接结论，并且与 M 的凸性无关. 现在我们证明关于相对内部运算的公式. 首先注意,

$$\mathrm{cl}\,(A(\mathrm{ri}\,M)) \supset A(\mathrm{cl}\,(\mathrm{ri}\,M)) = A(\mathrm{cl}\,M) \supset AM \supset A(\mathrm{ri}\,M).$$

这表明 AM 与 $A(\mathrm{ri}\,M)$ 有相同的闭包，从而也有相同的相对内部. 因此 $\mathrm{ri}\,(AM) \subset A(\mathrm{ri}\,M)$. 现在假定 $z \in A(\mathrm{ri}\,M)$, 并任取一点 $x \in AM$. 于是存在 $x' \in M$ 和 $z' \in \mathrm{ri}\,M$ 使得 $x = Ax', z = Az'$. 依据定理 1.4.11, 存在 $\mu > 1$ 使得 $(1-\mu)x' + \mu z' \in M$. 但 $A[(1-\mu)x' + \mu z'] = (1-\mu)x + \mu z \in AM$, 因此再应用定理 1.4.8 即得 $z \in \mathrm{ri}\,(AM)$. ∎

推论 1.4.15 对于任意凸集 $M \subset \mathbb{R}^n$ 和实数 λ, 有 $\mathrm{ri}\,(\lambda M) = \lambda \mathrm{ri}\,M$.

证明 留作练习. ∎

设 $M_1 \subset \mathbb{R}^n$ 和 $M_2 \subset \mathbb{R}^m$ 为两个凸集. 积集 $M_1 \times M_2$ 是指 $\mathbb{R}^{n+m}$ 中的子集:

$$M_1 \times M_2 = \{(x_1, x_2) \in \mathbb{R}^{n+m} \mid x_1 \in M_1, x_2 \in M_2\}.$$

显然我们有

$$\mathrm{ri}\,(M_1 \times M_2) = \mathrm{ri}\,M_1 \times \mathrm{ri}\,M_2,$$

$$\mathrm{cl}\,(M_1 \times M_2) = \mathrm{cl}\,M_1 \times \mathrm{cl}\,M_2.$$

据此并应用定理 1.4.14, 我们可得到如下推论:

推论 1.4.16 设 M_1 和 M_2 是 $\mathbb{R}^n$ 中的两个凸集. 那么

$$\mathrm{ri}\,(M_1 + M_2) = \mathrm{ri}\,M_1 + \mathrm{ri}\,M_2,$$

$$\mathrm{cl}\,(M_1+M_2)\supset \mathrm{cl}\,M_1+\mathrm{cl}\,M_2.$$

证明 定义线性变换 $A:\mathbb{R}^{2n}\to\mathbb{R}^n$ 为

$$A(x_1,x_2)=x_1+x_2,\quad \forall\, x_1,x_2\in\mathbb{R}^n,$$

我们有 $A(M_1\times M_2)=M_1+M_2$. ∎

定理 1.4.17 设 $A:\mathbb{R}^n\to\mathbb{R}^m$ 为线性变换. 设 M 为 $\mathbb{R}^m$ 中的凸集, 满足 $A^{-1}(\mathrm{ri}\,M)\neq\emptyset$. 那么,

$$\mathrm{ri}\,(A^{-1}M)=A^{-1}(\mathrm{ri}\,M),\quad \mathrm{cl}\,(A^{-1}M)=A^{-1}(\mathrm{cl}\,M).$$

证明 令 $N=\mathbb{R}^n\times M$, 并设 G 为 A 的图象, 即 $G=\{(x,Ax)\mid x\in\mathbb{R}^n\}$. 于是 G 是 $\mathbb{R}^n\times\mathbb{R}^m$ 中的一仿射集 (实际上是子空间), 并且 $G\cap\mathrm{ri}\,N\neq\emptyset$. 设 P 表示 $\mathbb{R}^n\times\mathbb{R}^m$ 到 $\mathbb{R}^n$ 的投影: $P(x,y)=x,\forall\, x\in\mathbb{R}^n,y\in\mathbb{R}^m$. 于是 $A^{-1}M=P(G\cap N)$. 应用定理 1.4.11, 可得

$$\mathrm{ri}\,(A^{-1}M)=P(\mathrm{ri}\,(G\cap N))=P(G\cap\mathrm{ri}\,N)=A^{-1}(\mathrm{ri}\,M),$$

$$\mathrm{cl}\,(A^{-1}M)=P(\mathrm{cl}\,(G\cap N))=P(G\cap\mathrm{cl}\,N)=A^{-1}(\mathrm{cl}\,M).$$

最后, 包含关系 $\mathrm{cl}\,(A^{-1}M)\subset A^{-1}(\mathrm{cl}\,M)$ 是 A 的连续性的结果. ∎

定理 1.4.18 设 M 为 $\mathbb{R}^n\times\mathbb{R}^m$ 中的凸子集. 对于每一 $y\in\mathbb{R}^m$, 设 $M_y=\{x\in\mathbb{R}^n\mid(x,y)\in M\}$. 记 $N=\{y\in\mathbb{R}^m\mid M_y\neq\emptyset\}$, 那么,

$$\mathrm{ri}\,M=\{(x,y)\mid y\in\mathrm{ri}\,N,x\in\mathrm{ri}\,M_y\}.$$

证明 定义投影 $P:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^m$,

$$P(x,y)=y,\quad \forall\,(x,y)\in\mathbb{R}^m\times\mathbb{R}^m.$$

显然, $PM=N$, 于是依据定理 1.4.14, $P(\mathrm{ri}\,M)=\mathrm{ri}\,N$,

这表明 $(x,y)\in \mathrm{ri}\, M \Longrightarrow y\in \mathrm{ri}\, N$. 现在设 $y\in \mathrm{ri}\, N$, 并作集合 $G_y=\{(x,y)\in I\!R^n\times I\!R^m \mid x\in I\!R^n\}$. 再依据定理 1.4.13, 我们有

$$G_y\cap \mathrm{ri}\, M=\mathrm{ri}\,(G_y\cap M)=\{(x,y)\mid x\in \mathrm{ri}\, M_y\}.$$

这样我们证明了，对于任意给定的 $y\in \mathrm{ri}\, N$,

$$(x,y)\in \mathrm{ri}\, M \Longleftrightarrow x\in \mathrm{ri}\, G_y.$$

定理证毕. ▮

定理 1.4.19 设 $M_1,\cdots,M_m$ 是 $I\!R^n$ 中的凸集，并设 $M_0=\mathrm{conv}\,(M_1\cup\cdots\cup M_m)$. 那么,

$$\mathrm{ri}\, M_0=\cup\left\{\sum_{k=1}^{m}\lambda_k \mathrm{ri}\, M_k \mid \lambda_i>0,\lambda_1+\cdots+\lambda_m=1\right\}.$$

证明 设 K_i 表示 $I\!R^{n+1}$ 中由 $\{(x_i,1)\mid x_i\in M_i\}$ 所生成的凸锥， $i=0,1,\cdots,m$, 那么不难验证

$$K_0=\mathrm{conv}\,(K_1\cup\cdots\cup K_m)=K_1+\cdots+K_m.$$

从而由推论 1.4.16 得到

$$\mathrm{ri}\, K_0=\mathrm{ri}\, K_1+\cdots+\mathrm{ri}\, K_m.$$

但是从定理 1.4.18 不难看出

$$\mathrm{ri}\, K_i=\{(x_i,\lambda_i)\mid \lambda_i>0, x_i\in \lambda_i \mathrm{ri}\, M_i\}.$$

因此

$$x_0\in \mathrm{ri}\, M_0 \Longleftrightarrow (x_0,1)\in K_0 \Longleftrightarrow$$

$$x_0\in \Big(\lambda_1 \mathrm{ri}\, M_i+\cdots+\lambda_m \mathrm{ri}\, M_m\Big),$$

这里 $\lambda_i>0,\lambda_1+\cdots+\lambda_m=1$. ▮

习 题 1.4

1.4.1 设 $x, y \in \mathbb{R}^n$, $[x,y] = \{\lambda x + (1-\lambda)y \mid 0 \le \lambda \le 1\}$. 试证 $\mathrm{ri}\,[x,y] = \{\lambda x + (1-\lambda)y \mid 0 < \lambda < 1\}$.

1.4.2 指出下列结论正确与否，如不正确，请举出反例：设 A 和 B 为 $\mathbb{R}^n$ 中的非空凸子集，如果 $A \subset B$, 则 $\mathrm{ri}\,A \subset \mathrm{ri}\,B$.

1.4.3 设 M 是 $\mathbb{R}^n$ 中的闭凸集. 试证：如果 x 和 y 是 M 的边界点，则或者 $[x,y] \subset \mathrm{bd}\,M$, 或者 $\mathrm{ri}\,[x,y] \subset \mathrm{int}\,M$.

1.4.4 设 M 是 $\mathbb{R}^n$ 中的凸子集，并且 $\mathrm{int}\,M \ne \emptyset$. 指出下列两个结论正确与否，如不正确，请举出反例：

(1) M 的边界是凸集；

(2) M 的边界不是凸集.

1.4.5 设 M 是 $\mathbb{R}^n$ 的凸子集，如果 $\mathrm{int}\,M \ne \emptyset$, 并且 M 还是紧集，那么 M 的边界不是凸集.

1.4.6 设 M 是 $\mathbb{R}^n$ 的凸子集，并设 F 是 $\mathbb{R}^n$ 的仿射集. 假定 $F \cap \mathrm{ri}\,M \ne \emptyset$, 试证：

$$\mathrm{ri}\,(F \cap M) = F \cap \mathrm{ri}\,M, \quad \mathrm{cl}\,(F \cap M) = F \cap \mathrm{cl}\,M.$$

1.4.7 设 M_1 和 M_2 是 $\mathbb{R}^n$ 的两个凸集，假定 $M_2 \subset \mathrm{cl}\,M_1$ 但 M_2 并不完全包含在 M_1 的相对边界中. 试证：$\mathrm{ri}\,M_2 \subset \mathrm{ri}\,M_1$.

1.4.8 设 M 是 $\mathbb{R}^n$ 中的非空凸集，并设 K 是 $\mathbb{R}^{n+1}$ 中由 $\{(x,1) \mid x \in M\}$ 所生成的凸锥. 试证：

$$\mathrm{ri}\,K = \{(x,\lambda) \in \mathbb{R}^{n+1} \mid \lambda > 0, x \in \lambda \mathrm{ri}\,M\}.$$

1.4.9 设 M 是 $\mathbb{R}^n$ 中的非空凸集，K 为由 M 生成的凸锥. 试证：

$$\mathrm{ri}\,K = \{\lambda x \mid \lambda > 0, x \in \mathrm{ri}\,M\}.$$

1.4.10 完成推论 1.4.6 和 1.4.7 的证明.

1.4.11 设 $M_1, \cdots, M_m$ 是 $\mathbb{R}^n$ 中的凸子集，试证：

$$\operatorname{conv}(M_1 \cup \cdots \cup M_m)$$
$$= \left\{ \sum_{k=1}^{m} \lambda_k x_k \;\middle|\; x_k \in M_k, \lambda_k \geq 0, \sum_{k=1}^{m} \lambda_k = 1 \right\}.$$

写出任意一族凸集的凸包的表达形式.

1.4.12 设 M 是 $\mathbb{R}^n$ 中的闭凸集，并设 $y \in \operatorname{int} M$, $x \notin M$, 试证 $\operatorname{ri}[x, y] \cap \operatorname{bd} M$ 是单点集.

1.5 凸集分离定理

凸集的分离是凸分析中最重要的概念之一，它所基于的基本事实是：$\mathbb{R}^n$ 中的一个超平面正好把 $\mathbb{R}^n$ 一分为二，并且此超平面的补集恰好是与其关联的两个不相交的开凸集(即两个开半空间) 之并.

我们知道，$\mathbb{R}^n$ 上任一线性泛函 f 对应于 $\mathbb{R}^n$ 中唯一点 x^* 使得

$$f(x) = \langle x, x^* \rangle, \quad \forall x \in \mathbb{R}^n.$$

通常我们把 f 与 x^* 等同. $\mathbb{R}^n$ 上的线性泛函全体构成的线性空间叫做 $\mathbb{R}^n$ 的对偶空间. 于是 $\mathbb{R}^n$ 的对偶空间等同于 $\mathbb{R}^n$ 自身. 注意 $\mathbb{R}^n$ 上的每一线性泛函都是连续的. 由 1.1 节我们知道，$\mathbb{R}^n$ 中任一超平面 H 都可以表示成

$$H(x^*, \beta) \triangleq \{x \in \mathbb{R}^n \mid \langle x, x^* \rangle = \beta\},$$

其中 $x^* \in \mathbb{R}^n$ 为非零元，$\beta \in \mathbb{R}$; 而且对于 $\lambda \neq 0$,

$$H(\lambda x^*, \lambda\beta) = H(x^*, \beta).$$

其逆命题也成立.

定理 1.5.1 设 $f,g\in\mathbb{R}^n$, $\alpha,\beta\in\mathbb{R}$, 并且 $H(f,\alpha)=H(g,\beta)$. 那么存在实数 $\lambda\neq 0$, 使得 $g=\lambda f, \beta=\lambda\alpha$.

证明 记 $H=H(f,\alpha)=H(g,\beta)$. 如果 f 或 g 为零元，则结论是显然的. 因此假定 $f\neq 0, g\neq 0$. 如果 $\alpha=0$, 则 H 是子空间， $0\in H$, 从而 $\beta=0$. 类似地，根据 $\beta=0$ 也可推出 $\alpha=0$. 于是我们先考虑 $\alpha=\beta=0$ 的情形. 选取一点 $z\in\mathbb{R}^n\backslash H$ 使得 $g(z)\neq 0$, 并令 $\lambda\triangleq f(z)/g(z)$. 由于 H 是 $n-1$ 维子空间，任意向量 $x\in\mathbb{R}^n$ 有唯一分解 $x=y+\mu z$, 其中 $y\in H$, $\mu\in\mathbb{R}$. 考虑到 $f(y)=g(y)=0$, 我们有

$$
\begin{aligned}
f(x)-\lambda g(x) &= f(y)+\mu f(z)-\lambda[g(y)+\mu g(z)] \\
&= \mu[f(z)-\lambda g(z)]=0, \quad \forall x\in\mathbb{R}^n,
\end{aligned}
$$

即

$$f=\lambda g.$$

现在设 $\alpha\neq 0$, $\beta\neq 0$. 今证 $H(f,0)=H(g,0)$. 事实上，任取一点 $z\in H$, 则 $f(z)=\alpha$. 如果 $x\in\mathbb{R}^n$, $f(x)=0$, 则 $f(x+z)=f(z)=\alpha$. 从而 $x+z\in H=H(g,\beta)$, 即 $g(x+z)=\beta$. 但 $g(z)=\beta$, 故 $g(x)=0$. 类似地，从 $g(x)=0$ 也可推出 $f(x)=0$. 因此 $H(f,0)=H(g,0)$. 由上面已经证明了的 $\alpha=\beta=0$ 的结果，我们得到 $f=\lambda g$, 其中 $\lambda=f(z)/g(z)=\alpha/\beta$. ∎

给定 $\mathbb{R}^n$ 中的子集 A 和超平面 $H(f,\alpha)$, 我们称集合 A 位于超平面 $H(f,\alpha)$ 的一侧，是指

$$f(x)\leq\alpha, \quad \forall x\in A,$$

或者

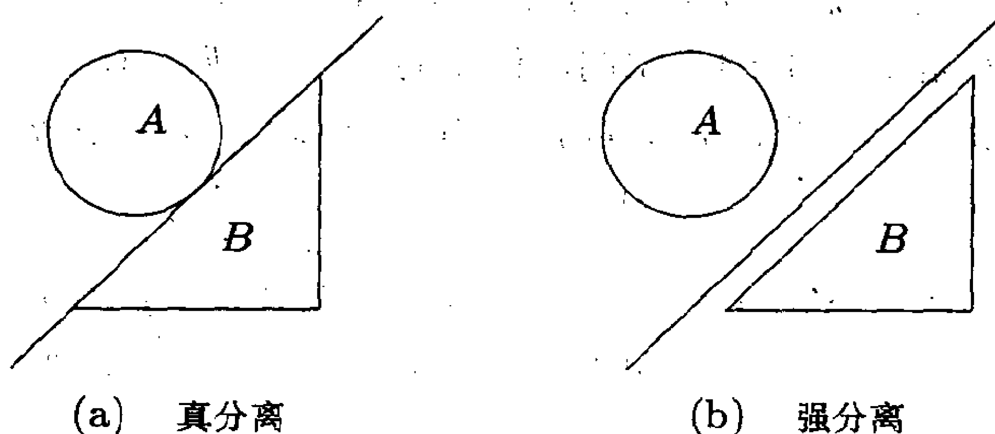
$$f(x)\geq\alpha, \quad \forall x\in A.$$

给定 $I\!R^n$ 中两个子集 A 和 B 以及一个超平面 $H=H(f,\alpha)$. 我们称 A 和 B 被超平面 H 所分离，是指它们分别位于 H 的两侧. 确切地说，就是

$$f(x)\le\alpha\le f(y),\quad \forall x\in A,\forall y\in B, \tag{1.5.1}$$

或者

$$f(x)\ge\alpha\ge f(y),\quad \forall x\in A,\forall y\in B. \tag{1.5.1$'$}$$



(a)　真分离　　　　(b)　强分离

A　　B

(c)　严格分离

图 1.5.1　凸集用超平面的分离

这样定义的集合的分离似乎并不令人满意，因为，例如在三维空间中位于同一平面中的两个集合按照这个定义已经算是分离了！因此有必要提出更严格的分离概念.（见图 1 5.1)

于是我们称 $I\!R^n$ 中的集合 A 和 B 被超平面 $H(f,\alpha)$ 真

分离，是指式 (1.5.1) 或者式 (1.5.1)′ 成立，同时式 (1.5.1) 或者式 (1.5.1)′ 不可能总是成立等号；称集合 A 和 B 被超平面 $H(f,\alpha)$ 严格分离，是指

$$f(x) < \alpha < f(y), \quad \forall x \in A, \forall y \in B, \tag{1.5.2}$$

或者

$$f(x) > \alpha > f(y), \quad \forall x \in A, \forall y \in B; \tag{1.5.2$'$}$$

而称集合 A 和 B 被超平面 $H(f,\alpha)$ 强分离，是指存在一实数 $\varepsilon > 0$, 使得

$$f(x) \leq \alpha - \varepsilon < \alpha + \varepsilon \leq f(y), \quad \forall x \in A, \forall y \in B, \tag{1.5.3}$$

或者

$$f(y) \leq \alpha - \varepsilon < \alpha + \varepsilon \leq f(x), \quad \forall x \in A, \forall y \in B. \tag{1.5.3$'$}$$

引理 1.5.2 设 A 和 B 为 $\mathbb{R}^n$ 中的两个非空子集. 那么 A, B 能用一个超平面分离的充分必要条件是存在一非零向量 $f \in \mathbb{R}^n$, 使得

$$\inf\{\langle x, f\rangle \mid x \in A\} \geq \sup\{\langle y, f\rangle \mid x \in B\}; \tag{1.5.4}$$

A, B 能用一个超平面真分离的充分必要条件是存在一非零向量 $f \in \mathbb{R}^n$ 满足式 (1.5.4) 及

$$\sup\{\langle x, f\rangle \mid x \in A\} > \inf\{\langle y, f\rangle \mid x \in B\}; \tag{1.5.5}$$

而 A, B 能用一个超平面强分离的充分必要条件是存在一非零向量 $f \in \mathbb{R}^n$ 满足

$$\inf\{\langle x, f\rangle \mid x \in A\} > \sup\{\langle y, f\rangle \mid x \in B\}. \tag{1.5.6}$$

证明 留作练习. ∎

这里要注意，真分离允许一个集合 (而非两个集合同时) 包含在分离超平面中. 例如，

$$A = \{(\xi, \eta) \in \mathbb{R}^2 \mid \xi > 0, \xi\eta \geq 1\},$$

$$B = \{(\xi, 0) \in \mathbb{R}^2 \mid \xi \geq 0\},$$

则 $A \cap B = \emptyset$, 并且能够分离这两个集合的唯一的超平面是 ξ 轴，它包含集合 B.

平面上两个不相交的三角形总能用一条直线 (即 $\mathbb{R}^2$ 中的超平面) 把它们分离开. 这个结论实际上对于 $\mathbb{R}^n$ 中的任意两个不相交的凸集都是对的. 在凸分析理论中，这一类结果统称为凸集分离定理，是凸集理论的重要的分析工具.

引理 1.5.3 设 A 是平面 $\mathbb{R}^2$ 中的非空相对开凸子集, $x_0 \notin A$, 那么存在通过点 x_0 的直线 L 使得 $A \cap A = \emptyset$.

证明 不失一般性，可以假定 $x_0 = 0$, 并且只要考虑 A 是 $\mathbb{R}^2$ 中开凸集的情形，因为当 A 是 $\mathbb{R}^2$ 中一维相对开凸集，即开线段时，结论是显然的. 于是设 A 是 $\mathbb{R}^2$ 中的开凸集， $0 \notin A$. 令

$$S = \{\lambda x \mid x \in A, \lambda > 0\},$$

则 S 是 $\mathbb{R}^2$ 中的一个开凸锥，它至多是一个开半平面. 任取 S 的一个非零边界点 x_1, 则通过 0 和 x_1 的直线 L 与 S 不相交，从而它也与 A 不相交 (见图 1.5.2). ∎

设 M 是 $\mathbb{R}^n$ 的一个子空间，则 $\mathbb{R}^n$ 有正交分解 $\mathbb{R}^n = M + M^{\perp}$, 这里 $M^{\perp}$ 是 M 的正交补. 于是任意元 $x \in \mathbb{R}^n$ 可唯一地分解成 $x = y + z$, 式中 $y \in M$, $z \in M^{\perp}$. 对应于这样的分解，我们定义线性映射 $P : \mathbb{R}^n \to M^{\perp}$,

$$Px = z. \tag{1.5.7}$$

P 称为 R^n 到 $M^{\perp}$ 的正交投影 (映射). 显然， P 是连续的，并且如果 A 是 $\mathbb{R}^n$ 的相对开集，则 A 的象 $P(A)$ 也是

相对开的 (见定理 1.4.11).

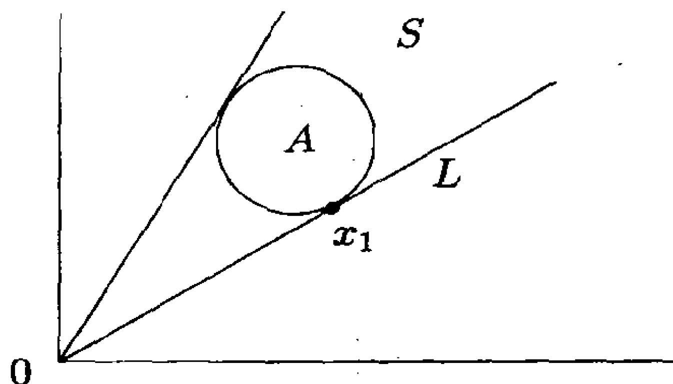


图 1.5.2

定理 1.5.4 设 A 是 $\mathbb{R}^n$ 中的一个非空相对开凸子集, M 是一个非空仿射集, 满足 $A \cap M = \emptyset$. 那么必定存在一个超平面 $H(f,\alpha)$ 使得 $M \subset H(f,\alpha)$, 并且 A 位于与超平面 $H(f,\alpha)$ 相关联的一个开半空间, 即

$$f(x) < \alpha, \quad \forall x \in A, \tag{1.5.8}$$

或者

$$f(x) > \alpha, \quad \forall x \in A. \tag{1.5.8$'$}$$

证明 如果 M 本身已经是一个超平面 $H(f,\alpha)$, 则由于 $A \cap M = \emptyset$, 式 (1.5.8) 或者式 (1.5.8)$'$ 必定成立. 因为否则的话, 存在两点 $x, y \in A$ 使得 $f(x) > \alpha$, 而 $f(y) < \alpha$. 令

$$\lambda = (\alpha - f(y)/(f(x) - f(y)),$$

则 $0 < \lambda < 1$. 由于 A 是凸集, $\lambda x + (1-\lambda)y \in A$. 但容易算出 $f(\lambda x + (1-\lambda)y) = \alpha$, 即 $\lambda x + (1-\lambda)y \in M$, 从而与 $M \cap A = \emptyset$ 相矛盾.

现在设 M 不是超平面, 从而 $\dim M = k < n-1$. 我们采用构造性的证明方法: 首先设法构造一个比 M 高一维

的含 M 的仿射子集 M_1, 并且 M_1 仍然与 A 不相交；如果 $\dim M_1 < n-1$, 则再构造一个比 M_1 高一维的含 M_1 的仿射子集 M_2, 并且 $M_2 \cap A = \emptyset$. 这一过程可以一直继续下去，直至构造出的仿射子集 H 的维数等于 $n-1$, 这时它正好是一个超平面，使得 $H \cap A = \emptyset$, 并且 $M \subset H$. 这个 H 即是我们所要找的超平面.

如果必要的话作一平移，不妨假定 $0 \in M$, 即 M 是一个线性子空间. 作 M 的正交补 $M^{\perp}$, 于是 $\mathbb{R}^n$ 可正交分解为 $\mathbb{R}^n = M + M^{\perp}$. 由于 $\dim M = k < n-1$, $M^{\perp}$ 包含一个二维子空间 N. 设 $P: \mathbb{R}^n \to M^{\perp}$ 是 $\mathbb{R}^n$ 到 $M^{\perp}$ 上的正交投影. 所以 $P(A) \subset M^{\perp}$. 记

$$A_1 \triangleq N \cap P(A).$$

A_1 是平面 N 中的相对开凸集. 我们有 $0 \notin A_1$, 因为否则的话， $A \cap M \neq \emptyset$! 于是根据引理 1.5.3, 存在 N 中的一维子空间 L (通过 0 点的直线), 使得 L 与 A_1 不相交，从而 $L \cap P(A) = \emptyset$. 现在令 $M_1 = P^{-1}(L)$, 则 M_1 是一个 $k+1$ 维子空间，并且满足 $M_1 \supset M$, $M_1 \cap A = \emptyset$. 事实上，设 $L = \{\lambda z \mid \alpha \in \mathbb{R}\}$, 并设 $y \in M^{\perp}$ 使得 $Py = z$, 于是 $M_1 = \{\alpha y + x \mid \alpha \in \mathbb{R}, x \in M\}$. 由此可见, $\dim M_1 = k+1$; $M_1 \supset M$ 是显然的；如果 $M_1 \cap A \neq \emptyset$, 则存在 $\alpha \in \mathbb{R}$ 和 $x \in M$ 使得 $\alpha y + x \in A$. 但这样我们就得出 $\alpha z = P(\alpha y + x) \in L \cap P(A)$, 与 $L \cap P(A) = \emptyset$ 相抵触. ∎

定理 1.5.5 设 A_1 和 A_2 是 $\mathbb{R}^n$ 中的两个非空凸子集，那么为了 A_1 和 A_2 能用一超平面把它真分离，必须且只须 $\operatorname{ri} A_1 \cap \operatorname{ri} A_2 = \emptyset$.

证明 令 $A = A_1 - A_2$, 则 A 也是 $\mathbb{R}^n$ 中的凸集, 并且依据推论 1.4.12, $\operatorname{ri} A = \operatorname{ri} A_1 - \operatorname{ri} A_2$. 于是

$$0 \notin \operatorname{ri} A \Longleftrightarrow \operatorname{ri} A_1 \cap \operatorname{ri} A_2 = \emptyset.$$

今若 $\operatorname{ri} A_1 \cap \operatorname{ri} A_2 = \emptyset$, 即 $0 \notin \operatorname{ri} A$, 则依据定理 1.5.4, 存在一个超平面 H 包含 $M = \{0\}$, 并且使得 $\operatorname{ri} A$ 位于一个与该超平面 H 相关联的半开空间内. 由于

$$A \subset \operatorname{cl}(\operatorname{ri} A) = \operatorname{cl} A,$$

A 必包含在相应的闭半空间内. 于是存在非零向量 $f \in \mathbb{R}^n$ 满足

$$0 \le \inf_{x \in A} \langle x, f \rangle = \inf_{x_1 \in A_1} \langle x_1, f \rangle - \sup_{x_2 \in A_2} \langle x_2, f \rangle,$$

$$0 < \sup_{x \in A} \langle x, f \rangle = \sup_{x_1 \in A_1} \langle x_1, f \rangle - \inf_{x_2 \in A_2} \langle x_2, f \rangle.$$

但依据引理 1.5.2, 这正是说, A_1 和 A_2 被超平面 $H(f, \alpha)$ 真分离, 其中 α 满足

$$\inf_{x_2 \in A_2} \langle x, f \rangle \le \alpha \le \sup_{x_1 \in A_1} \langle x_1, f \rangle.$$

反过来, 如果 A_1 和 A_2 被某个超平面 $H(f, \alpha)$ 真分离, 于是, 例如设 A_1 真在 $H(f, \alpha)$ 的一侧, 即存在点 $x \in A_1$ 使得 $f(x) > \alpha$. 注意如果 $A_1 \cap H(f, \alpha) = \emptyset$, 则 $A_1 \cap A_2 = \emptyset$. 因此只需考虑 $A_1 \cap H(f, \alpha) \neq \emptyset$ 的情形. 任取一点 $z \in A_1 \cap H(f, \alpha)$, 今证 $z \notin \operatorname{ri} A_1$. 事实上, 如果 $z \in \operatorname{ri} A_1$, 则由定理 1.4.11, 存在 $\mu > 1$ 使得 $(1 - \mu)x + \mu z \in A_1$, 即 $f((1 - \mu)x + \mu z) \ge \alpha$. 但 $f(x) > \alpha$, 而且 $f(z) = \alpha$, 故

$$f((1 - \mu)x + \mu z) = (1 - \mu) f(x) + \mu \alpha < \alpha,$$

得出矛盾. 这样我们证明了 $\operatorname{ri} A_1 \cap H(f, \alpha) = \emptyset$, 从而 $A_2 \cap \operatorname{ri} A_1 = \emptyset$. ∎

注 (1) 从定理的证明看出，条件 $\mathrm{ri}\,A_1 \cap \mathrm{ri}\,A_2 = \emptyset$ 可以用 $\mathrm{ri}\,A_1 \cap A_2 = \emptyset$ 或者 $A_1 \cap \mathrm{ri}\,A_2 = \emptyset$ 代替. 因此，$\mathbb{R}^n$ 中两个不相交的凸集可以用一个超平面真分离，这与人们在二、三维空间中的直觉是一致的. 但在无穷维空间，即使是无穷维的 Hilbert 空间，为了分离两个凸集 A_1 和 A_2，上述条件是不够的，必须用更强的条件 $A_1 \cap \mathrm{int}\,A_2 = \emptyset$ 或 $A_2 \cap \mathrm{int}\,A_1 = \emptyset$ 代替. 这已经超出本书讨论的范围了.

(2) 如果一个凸集 $A \subset \mathbb{R}^n$ 位于超平面 $H(f,\alpha)$ 的一侧，并且又不包含在 $H(f,\alpha)$ 中，那么 $\mathrm{ri}\,A$ 必定严格位于 $H(f,\alpha)$ 的一侧；确切地说，例如有

$$f(x) \le \alpha, \quad \forall\, x \in A,$$

$$f(x) < \alpha, \quad \forall\, x \in \mathrm{ri}\,A.$$

当然这里的不等号可全部反向.

为了两个凸集能用一个超平面强分离，必须要加上更强的条件. 容易给出 $\mathbb{R}^2$ 中的例子，即使不相交的两个闭凸集也未必能用一个超平面强分离，见图 1.5.1(c).

定理 1.5.6 设 M_1 和 M_2 是 $\mathbb{R}^n$ 中的两个非空凸子集. 那么为了 M_1 和 M_2 能被一个超平面将它们强分离，必须且只须

$$d(M_1, M_2) \triangleq \inf\{\|x_1 - x_2\| \mid x_1 \in M_1, x_2 \in M_2\} > 0, \tag{1.5.9}$$

换句话说，$0 \notin \mathrm{cl}\,(M_1 - M_2)$.

证明 如果 M_1 和 M_2 能被超平面 $H(f,\alpha)$ 强分离，则依据定义，存在 $\varepsilon > 0$, 使得

$$f(x) \le \alpha - \varepsilon < \alpha + \varepsilon \le f(y), \quad \forall\, x \in M_1, \forall\, y \in M_2.$$

于是

$$\|f\|\,\|x-y\| \geq f(y-x) \geq 2\varepsilon, \quad \forall\, x \in M_1, \forall\, y \in M_2,$$

即

$$\|x-y\| \geq 2\varepsilon/\|f\|, \quad \forall\, x \in M_1, \forall\, y \in M_2,$$

从而式 (1.5.9) 成立. 反之, 设式 (1.5.9) 成立. 设 $2\delta \triangleq d(M_1, M_2)$, 并令

$$M_1^\delta \triangleq M_1 + B(0,\delta) \quad M_2^\delta \triangleq M_2 + B(0,\delta).$$

显然 M_1^δ 和 M_2^δ 是 $\mathbb{R}^n$ 中的开凸集, 并且 $M_1^\delta \cap M_2^\delta = \emptyset$. 从而由定理 1.5.5, 存在非零元 $f \in \mathbb{R}^n$ 和实数 $\alpha \in \mathbb{R}$ 使得

$$f(x) \leq \alpha \leq f(y), \quad \forall\, x \in M_1^\delta, \forall\, y \in M_2^\delta.$$

由于 $f \neq 0$, 存在 $u_0 \in B(0,\delta)$ 使得 $f(u_0) > 0$. 这样, 如果令 $\varepsilon = f(u_0)$, 则得

$$f(x_1) \leq \alpha - \varepsilon < \alpha + \varepsilon \leq f(x_2), \quad \forall\, x_1 \in M_1, \forall\, x_2 \in M_2,$$

即 M_1 和 M_2 被超平面 $H(f,\alpha)$ 强分离. ∎

推论 1.5.7 设 M_1 和 M_2 是 $\mathbb{R}^n$ 中的两个不相交的闭凸子集. 如果其中之一为有界集, 则必定存在一个超平面把它们强分离.

证明 只需注意在推论的假设下, $d(M_1, M_2) > 0$. ∎

设 I 为某一指标集, 给定 $b_i \in \mathbb{R}^n$, $i \in I$. 我们知道线性不等式组

$$\langle x, b_i \rangle \leq \beta_i, \quad i \in I \tag{1.5.10}$$

的解 x 的集合正好是一族闭半空间之交, 从而是一个闭凸子集. 下面我们指出, $\mathbb{R}^n$ 中的每一个闭凸子集都可以表示成某个不等式组的解集.

定理 1.5.8 $\mathbb{R}^n$ 中的闭凸子集 M 是一切包含 M 的闭半空间之交.

证明 不妨假定 $\emptyset \neq M \neq \mathbb{R}^n$. 给定任意 $a \notin M$, 集合 $M_1 = \{a\}$ 和 $M_2 = M$ 满足定理 1.5.6 或推论 1.5.7 的条件, 因此存在一个超平面把它们强分离. 与此超平面相关联的一个闭半空间包含 M 但不包含点 a. 因此, 如果记 M' 为一切包含 M 的闭半空间之交, 则 $a \notin M'$. 由于 $a \notin M$ 的任意性, 即得 $M' = M$. ▮

推论 1.5.9 设 M 是 $\mathbb{R}^n$ 中的真凸子集 (即 $M \neq \mathbb{R}^n$), 那么存在一个闭半空间包含 M.

证明 留作练习. ▮

习 题 1.5

1.5.1 完成引理 1.5.2 的证明.

1.5.2 给定集合 $M \subset \mathbb{R}^n$ 和超平面 $H(f,\alpha)$, 称集合 M 为被 $H(f,\alpha)$ 所切割, 是指 M 不位于 $H(f,\alpha)$ 的一侧, 即存在两点 $x, y \in M$, 使得 $f(x) < \alpha$, 而 $f(y) > \alpha$. 试证: 超平面 $H(f,\alpha)$ 切割集合 M 的充分必要条件是:

(1) $H(f,\alpha)$ 不包含 M;

(2) $H(f,\alpha) \cap \operatorname{ri} M \neq \emptyset$.

1.5.3 设 A 和 B 是 $\mathbb{R}^n$ 中的两个非空紧子集. 试证: 为了存在一个超平面把 A 和 B 强分离, 必须且只须 $\operatorname{conv} A \cap \operatorname{conv} B = \emptyset$.

1.5.4 完成推论 1.5.9 的证明.

1.5.5 设 A 和 B 是 $\mathbb{R}^n$ 中的两个非空紧子集. 试证: 为了存在一个超平面把 A 和 B 强分离, 必须且只须对于 B 的每个至多含 $n+1$ 点的子集 M, 都能用一个超平面把 M 和 A 强分离.

1.5.6 设 M 是 $\mathbb{R}^n$ 中的真凸子集, 试证: $\operatorname{cl} M$ 也是 $\mathbb{R}^n$ 中的真凸子集.

1.5.7 设 A 和 B 是 $\mathbb{R}^n$ 的两个不相交的凸子集, 并且 $A \cup B$

是仿射集. 试证：$A\cup B=\operatorname{aff}A=\operatorname{aff}B$, 并举例说明上述不相交和凸性两个条件是不可缺少的.

1.5.8 设 A 和 B 是 $\mathbb{R}^n$ 的两个不相交的开凸子集，试证：存在一个超平面把它们严格分离.

1.5.9 设 A 和 B 是 $\mathbb{R}^n$ 的两个凸子集，试证：如果 $\operatorname{int}B\neq\emptyset$ 并且 $A\subset\operatorname{bd}B$, 则 $\operatorname{aff}A\cap\operatorname{int}B=\emptyset$.

1.6 承托超平面

相切这一概念是分析中的一个重要工具. 曲线的切线，曲面的切平面等通常都是通过微分来定义的. 在凸分析中，则利用凸集的分离来定义广义相切. 凸集的广义相切通过所谓“承托超平面”来描述 (见图 1.6.1), 这一概念在今后将用以讨论凸函数的次微分. 承托超平面又称支撑超平面.

平面上通过凸集的每一边界点都可以作一条直线 (“承托”直线), 使得该凸集位于此“承托”直线的一侧. 其实这一事实对于 $\mathbb{R}^n$ 中的一般凸集也是对的，并且它可以作为凸集的另一种刻划方式.

一般地,对于 $\mathbb{R}^n$ 中的集合 M 和超平面 $H(f,\alpha)$, $H(f,\alpha)$ 称为 M 的承托超平面，是指 M 位于 $H(f,\alpha)$ 的同一侧，并且 M 与 $H(f,\alpha)$ 至少有一个交点；换句话说，$\langle x,f\rangle\leq\alpha,\forall x\in M$, 并且至少有一个点 $x_0\in M$ 使得 $\langle x_0,f\rangle=\alpha$. 这时 $H(f,\alpha)$ 叫做 M 的通过点 $x_0\in M$ 的承托超平面，相应的 f 叫做承托线性泛函； f 作为 $\mathbb{R}^n$ 的向量是该承托超平面 $H(f,\alpha)$ 的法向量，在 M 是凸集的情形下，通常也叫做 M 在 x_0 处的法向量.

注意，如果 $\dim M<n$, 则我们总能把 $\operatorname{aff}M$ 扩张成一

个包含整个 M 的超平面，它当然是在上述意义下的承托超平面. 但这样的承托超平面显然意义不大，因此，今后如无特别说明，当谈到承托超平面时，总是指 M 的非不足道的承托超平面，即 M 不整个地包含在此超平面内.

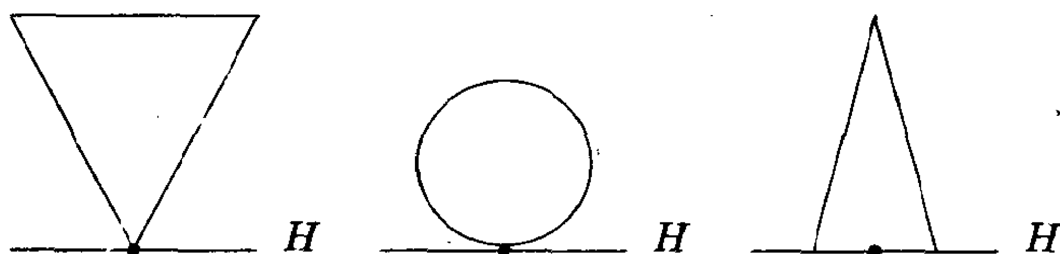


图 1.6.1　凸集的承托超平面

我们回忆一下，所谓集合 M 的相对边界 $\mathrm{bd_r}\,M$, 是指集合 $\mathrm{cl}\,M\backslash\mathrm{ri}\,M$.

定理 1.6.1　设 M 是 $\mathbb{R}^n$ 中的凸子集，N 是 M 的非空凸子集 (例如由单点组成的集合). 那么为了存在一个 M 的包含 N 的承托超平面，必须且只须 $N\cap(\mathrm{ri}\,M)=\emptyset$.

证明　注意超平面 $H(f,\alpha)$ 是 M 的包含 N 的承托超平面等价于 $H(f,\alpha)$ 把 M 与 N 真分离. 但依据定理 1.5.5 的注，后者的充分必要条件是 $N\cap(\mathrm{ri}\,M)=\emptyset$. ∎

推论 1.6.2　通过闭凸集 $M\subset\mathbb{R}^n$ 的每一个相对边界点都可以作一个 M 的承托超平面. ∎

定理 1.6.3　设 M 为 $\mathbb{R}^n$ 的闭子集，并且 $\mathrm{ri}\,M\neq\emptyset$. 如果通过 M 的每一个相对边界点都可以作 M 的一个承托超平面，则 M 是凸集.

证明　情况 $n=1$ 的证明是很容易的，留作练习. 现证明 $n\geq2$ 的情况. 如果 $M=\mathbb{R}^n$, 则 M 显然是凸的. 于是

假定 $M \neq \mathbb{R}^n$.

(1) $\operatorname{aff} M = \mathbb{R}^n$. 这时 $\operatorname{int} M \neq \emptyset$. 任取一点 $x \notin M$ 和一点 $y \in \operatorname{int} M$ (见图 1.6.2). 于是存在 M 的一边界点 b 使得 $b \in \operatorname{ri}[x, y]$ (这里 $[x, y]$ 为连接 x 和 y 的闭线段). 通过点 b 的 M 的承托超平面不包含点 x, 因为否则的话, 它要包含连接 x 和 y 的整个直线, 从而包含内点 y, 但这是不可能的. 这样我们证明了 M 等于一切包含 M 的闭半空间之交, 从而是凸集.

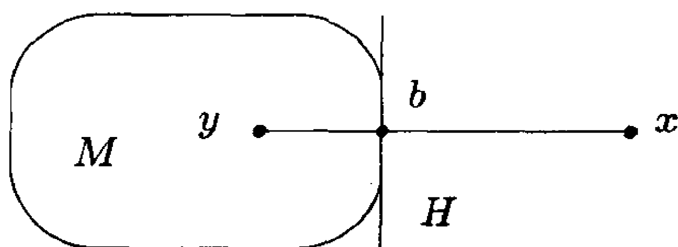


图 1.6.2

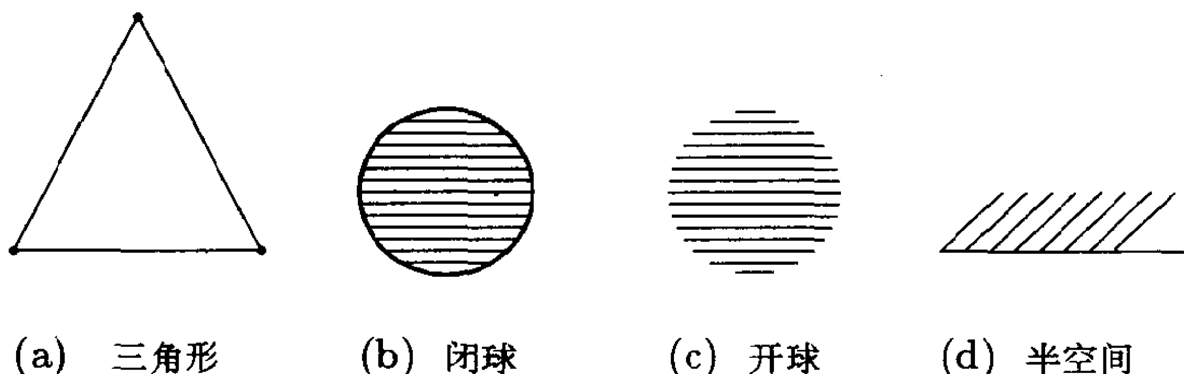
(2) $\operatorname{aff} M < n$. 如果 $x \notin \operatorname{aff} M$, 则依据定理 1.5.4, 存在一个包含 $\operatorname{aff} M$ 的超平面, 使得 $x \notin H$; 而如果 $x \in \operatorname{aff} M \backslash M$, 则由假设 $\operatorname{ri} M \neq \emptyset$, 利用与 (1) 相类似的方法可以证明存在一个超平面 H 使得 M 与 x 位于 H 的不同侧. 这样我们同样得出结论: M 等于一切包含 M 的闭半空间之交. ▌

推论 1.6.4 设 M 为 $\mathbb{R}^n$ 的闭子集, $\operatorname{ri} M \neq \emptyset$. 那么 M 是凸集的充分必要条件是: 通过其每一个相对边界点都可以作一个承托超平面.

下面我们介绍承托超平面概念的一个重要应用.

给定一个凸集 $M \subset \mathbb{R}^n$, 能否把 M 表示成其某个子

集 M' 的凸包？如果 M 由多于一个点组成，则答案是肯定的．例如我们可以取 $M' = M\backslash\{x\}$，其中 $x \in \operatorname{ri} M$．如果 M 还是一个紧集，则我们可以取 $M' = \operatorname{bd} M$．于是我们要问：一个紧凸集的边界是不是具有这种性质的最小的子集呢？观察一下三角形、矩形等图形就不难得出否定的结论．容易验证，它们的顶点集合才是具有这种性质的最小集合．分析一下这些顶点，我们发现它们的共同的特征是它们不可能是该集合上任何一根线段的内点．于是对于一般的凸集，我们给出如下的定义：



(a) 三角形　(b) 闭球　(c) 开球　(d) 半空间

图 1.6.3　凸集的极点

设 $M \subset \mathbb{R}^n$ 是一个凸集．我们称点 $x \in M$ 是 M 的一个极点，是指 x 不能表示成 M 中任何线段的中间点，即不存在 $y, z \in M$ 和 $\mu \in (0,1)$ 使得 $x = (1-\mu)y + \mu z$．M 的所有极点的全体记作 $E(M)$．

容易看出，三角形的极点集由其三个顶点组成；闭球的极点集是其边界；开球的极点集是空集；而闭半空间的极点集也是空集 (见图 1.6.3)

显然，为了凸集 M 中的点 x 是其极点，必须且只须 $M\backslash\{x\}$ 是凸集．由此可见，为了 M 的凸子集 M' 具有性质：

$\operatorname{conv} M' = M$, M' 必须包含 M 的极点集 $E(M)$. 从上面的例子可以看出，为了使得子集 $M' \subset M$ 满足 $\operatorname{conv} M' = M$, 一般说来，$M'$ 要比 $E(M)$ 大. 但是当 M 是紧凸集时，M' 可以取成 $E(M)$(定理 1.6.5). 这就是说，每个紧凸集 M 至少有一个极点并且其极点集 $E(M)$ 是使之凸包等于 M 的 M 中最小的子集.

定理 1.6.5 设 M 为 $\mathbb{R}^n$ 的紧凸集，那么 $\operatorname{conv} E(M) = M$, 并且每一点 $x \in M$ 可以表示为 $E(M)$ 中至多 $n+1$ 个点的凸组合.

证明 我们用数学归纳法来证明. 当 $n=1$ 时，紧凸集 M 是闭区间 $[a,b]$, 其极点集 $E(M) = \{a,b\}$. 从而定理的结论显然成立. 现在设 $n<k$ 时定理成立，今证 $n=k$ 时定理也成立. 设 $\dim M = p, x \in M$. 如果 $x \in \operatorname{bd} M \backslash \operatorname{ri} M$, 则依据推论 1.6.2, 通过点 x 可以作 M 的一个承托超平面 H. 由归纳假设，x 是 $M \cap H$ 的至多 p 个极点的凸组合. 但 $M \cap H$ 的极点也是 M 的极点 (见习题 1.6.3), 因此，x 是 M 的至多 p 个极点的凸组合.

再考虑 $x \in \operatorname{ri} M$ 的情况. 我们先假定 $E(M) \neq \emptyset$. 取一点 $y \in E(M)$, 并把线段 $[x,y]$(在 $\operatorname{aff} M$ 中) 向左方向延伸至 M 的相对边界点 z. 这样，存在 $\mu \in (0,1)$ 使得 $x = (1-\mu)y + \mu z$. 另一方面，由上述已经证明了的部分的结论，z 可以表示成 $E(M)$ 中至多 p 个点的凸组合，即 $z = \mu_1 y_1 + \cdots + \mu_p y_p$, 其中 $y_i \in E(M)$, $1 \le i \le p$. 由此可见，x 可以表示成 $E(M)$ 中至多 $p+1$ 个点的凸组合，这里 $p \le k$.

剩下证明 $E(M) \neq \emptyset$. 我们还是用数学归纳法来证明. $n=1$ 时显然有 $E(M) \neq \emptyset$. 现在设当 $n<p$ 时 $E(M) \neq \emptyset$.

今证 $n=p$ 时此结论仍成立. 设 H 是 M 的一个承托超平面, 它的存在性由推论 1.6.2 保证. 于是 $H\cap M$ 是一个维数至多为 $p-1$ 的紧凸集. 根据归纳假设, $M\cap H$ 至少有一个极点, 但此极点也是 M 的极点. 证毕. ▮

定理 1.6.6 设 M 为 $\mathbb{R}^n$ 的紧凸集, 又设 f 是 $\mathbb{R}^n$ 上的线性泛函. 那么存在 M 的极点 x_0 和 x_1, 使得

$$f(x_0)=\max\{f(x)\mid x\in M\},$$

$$f(x_1)=\min\{f(x)\mid x\in M\}$$

证明 由于 f 的连续性和 M 的紧性, f 在 M 上达到其最大值和最小值, 比方说, 点 $x',y'\in M$, 使得

$$f(x')=\max\{f(x)\mid x\in M\},$$

$$f(y')=\min\{f(x)\mid x\in M\}$$

由于 M 是紧凸集, 从定理 1.6.5, 存在 M 的极点 $x_1,\cdots,x_m$ 和非负数 $\lambda_1,\cdots,\lambda_m$

$$x'=\lambda_1x_1+\cdots+\lambda_mx_m,\ \ \lambda_1+\cdots+\lambda_m=1.$$

如果 M 的极点中没有一个达到最大值 $f(x')$, 则

$$f(x_k)<f(x'),\quad k=1,\cdots,m.$$

由此得出

$$f(x')=f\left(\sum_{k=1}^{m}\lambda_kx_k\right)\le\sum_{k=1}^{m}\lambda_kf(x_k)<f(x').$$

这不可能, 所以 f 必在 $E(M)$ 上达到最大值.

f 在 $E(M)$ 上达到最小值可以用类似的方法证明. ▮

习 题 1.6

1.6.1 就 $n=1$ 的情形证明定理 1.6.3.

1.6.2 试证: 为了凸集 M 中的点 x 是其极点, 必须且只须 $M\backslash\{x\}$ 是凸集.

1.6.3 设 M 是 $\mathbb{R}^n$ 中的闭凸子集, H 是 M 的通过点 $x\in M$ 的承托超平面. 试证: $x\in E(H\cap M)\iff x\in E(M)$.

1.6.4 找出 $\mathbb{R}^2$ 中的一个闭凸集 M, 使得 $E(M)\neq\emptyset$, 但 $\operatorname{conv}E(M)\neq M$.

1.6.5 构造 $\mathbb{R}^2$ 中的一个有界凸集 M, 使得 $E(M)\neq\emptyset$, 但 $\operatorname{conv}E(M)\neq M$.

1.6.6 设 M' 是凸集 $M\subset\mathbb{R}^n$ 的子集, 使得 $\operatorname{conv}M'=M$. 试证: $E(M)\subset M'$.

1.6.7 设 M 是 $\mathbb{R}^n$ 的一个有界凸集, 试证: $\operatorname{conv}(\operatorname{bd}M)=\operatorname{cl}M$, 并举例说明这里凸集的有界性是不可缺少的.

1.6.8 设 $M\subset\mathbb{R}^n$ 是紧凸集使得 $\operatorname{int}M\neq\emptyset$. 应用定理 1.6.5 证明: M 的边界不是凸的.

1.6.9 设 M 是 $\mathbb{R}^n$ 中一个至少含有两个点的闭真凸集. 试证: M 是凸锥的充分必要条件是: M 的所有承托超平面至少有一个公共点.

1.6.10 设 $M\subset\mathbb{R}^n$ 为闭凸集, $\operatorname{bd}M\neq\emptyset$. 设 f 为 $\mathbb{R}^n$ 上的线性泛函. 假定 f 在 M 上是上有界的, 即 $\sup\{f(x)\mid x\in M\}<\infty$, 试证:

$$\sup\{f(x)\mid x\in M\}=\sup\{f(x)\mid x\in\operatorname{bd}M\}.$$

1.7 凸多面体

在 1.3 节中, 我们把 $\mathbb{R}^n$ 中的有限点集的凸包称作凸

多面体，这里我们有时也简称作多面体. 这是平面上多边形和三维空间中多面体的自然推广. 这些低维空间中的多面体早在 2000 多年前就被研究过. 随着线性规划和博奕论的出现，在过去几十年中人们对高维空间中的多面体作了广泛深入的研究. 本节仅对多面体的理论作一简单介绍，有兴趣的读者可参阅 Grünbaum 的 "Convex Polytopes".

首先我们约定一些术语. 设 P 是 $\mathbb{R}^n$ 中任一多面体. P 的子集 F 叫做 P 的一个面，是指或者 $F=\emptyset$, 或者 $F=P$, 或者有 P 的一个承托超平面 H 使得 $F=P\cap H$. P 和 $\emptyset$ 称为 P 的非真面，而其余面则称为 P 的真面. 如果 F 的维数为 k, 则 F 称为 P 的 k 维面. 如果多面体 P 的维数为 k, 则 P 称为 k 维多面体.

习惯上我们把 P 的零维面叫做顶点，P 的所有顶点构成的集合记作 $V(P)$. P 的一维面叫做棱边，而 $n-1$ 维面叫做 P 的刻面 (facet).

$\mathbb{R}^n$ 中的有限集合 $\{x_1,\cdots,x_m\}$ 叫做多面体 P 的极小表现，是指它满足:

(1) $P=\operatorname{conv}\{x_1,\cdots,x_m\}$;

(2) $x_k\notin\operatorname{conv}\{x_1,\cdots,x_{k-1},x_{k+1},\cdots,x_m\}$, $1\le k\le m$.

不难看出，每一个多面体 P 都有极小表现，并且其极小表现是唯一的. 事实上，P 作为多面体总能表示为 $P=\operatorname{conv}\{x_1,\cdots,x_l\}$. 如果某个 x_j 是其余 x_i 的凸组合，则 x_j 可以从这组点中剔除，而使剩下的点集的凸包仍然是多面体 P. 这一过程可以继续下去，直至留下的点组成 P 的极小表现.

对于多面体 $P=\{x_1,\cdots,x_m\}$, 当 $\{x_1,\cdots,x_m\}$ 是 P

的极小表现时，人们自然指望 P 的顶点就是 P 的极点，并且正好是 $x_1,\cdots,x_m$.

定理 1.7.1 设 $S=\{x_1,\cdots,x_m\}$ 是多面体 P 的极小表现. 那么下列三个命题彼此等价：

(1) $x\in S$;

(2) x 是 P 的顶点；

(3) x 是 P 的极点.

证明 (1)$\Longrightarrow$(2). 假定 $x\in S$, 并设 $M=\operatorname{conv}(S\backslash\{x\})$. 从 S 的极小表现性质知道 $x\notin M$. 由于 M 是紧凸集，根据推论 1.5.7, 存在一个超平面 $H(f,\alpha)$ 把 $\{x\}$ 和 M 强分离. 不妨设 $x\in\operatorname{int}H^-(f,\alpha)$, $M\subset\operatorname{int}H^+(f,\alpha)$, 其中

$$H^-(f,\alpha)=\{z\in I\!R^n\mid f(z)\le\alpha\},$$

$$H^+(f,\alpha)=\{z\in I\!R^n\mid f(z)\ge\alpha\}.$$

于是 $f(x)<\alpha$. 令 $\beta=f(x)$. 超平面 $H(f,\beta)$ 与 $H(f,\alpha)$ 平行，并且 $x\in H(f,\beta)$ (见图 1.7.1). 此外，x 是多面体 P 中位于超平面 $H(f,\beta)$ 上的唯一点. 事实上，若还有点 $y\in H(f,\beta)\cap P$, 则 y 可表示成

$$y=\lambda+(1-\lambda)z,$$

其中 $0<\lambda\le1$, $z\in M$. 于是

$$\beta=f(y)=\lambda\beta+(1-\lambda)f(z)\ge\lambda\beta+(1-\lambda)\alpha,$$

即

$$(1-\lambda)\beta\ge(1-\lambda)\alpha.$$

由于 $\beta<\alpha$, 这只有当 $\lambda=1$ 时才有可能. 于是

$$H(f,\beta)\cap P=\{x\},$$

从而 x 是 P 的顶点.

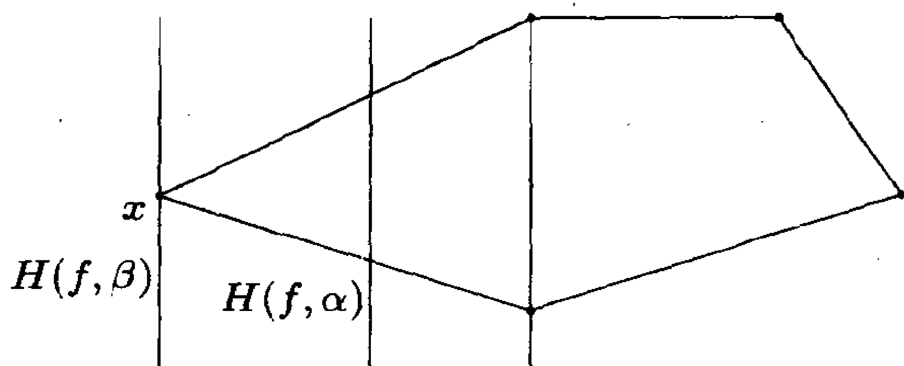


图 1.7.1

(2)$\Longrightarrow$(3) 是显然的. 最后证明 (3)$\Longrightarrow$(1). 设 x 是 P 的极点，作为 P 中的点，x 可以表示成 $x_1,\cdots,x_m$ 的凸组合，$x=\lambda_1x_1+\cdots+\lambda_mx_m$. 如果 $0\leq\lambda_i<1,\forall i=1,\cdots,m$，则不难看出，$x$ 也可以表示成 P 中某线段的相对内点，从而是 P 的相对内点，这不可能. 因此必存在一个标号 k 使得 $\lambda_k=1$, 从而 $x=x_k$. ▮

定理 1.7.2 设 P 为 $\mathbb{R}^n$ 中的一个多面体. 那么 P 的每一个真面本身也是一个多面体，并且 P 只有有穷多个不同的面.

证明 设 $\{x_1,\cdots,x_m\}$ 为 P 的极小表现，并假定 F 为 P 的一个真面，而 $H(f,\alpha)$ 为 P 的一承托超平面，使得 $F=H(f,\alpha)\cap P$. 不失一般性，我们可以假定

$$\{x_1,\cdots,x_k\}\subset H(f,\alpha);\ f(x_i)>\alpha,\ i=k+1,\cdots,m.$$

换句话说，

$$f(x_i)=\alpha,\quad i=1,\cdots,k,$$
$$f(x_j)=\alpha+\varepsilon_j,\quad j=k+1,\cdots,m,$$

其中 $\varepsilon_j>0,\ j=k+1,\cdots,m$. 现在设

$$x=\lambda_1x_1+\cdots+\lambda_mx_m\ \left(\lambda_i\geq0,\ \lambda_1+\cdots+\lambda_m=1\right)$$

是 P 的任意点. 那么

$$f(x) = \sum_{i=1}^{k} \lambda_i \alpha + \sum_{i=k+1}^{m} \lambda_i(\alpha + \varepsilon_i)$$
$$= \alpha + \sum_{i=k+1}^{m} \lambda_i \varepsilon_i.$$

于是 $x \in H(f,\alpha) \iff f(x) = \alpha \iff \sum_{i=1}^{k} \lambda_i \varepsilon_i = 0$. 但 $\lambda_i \geq 0$, $\varepsilon_i > 0$, 所以 $x \in H(f,\alpha) \iff \lambda_{k+1} = \cdots = \lambda_m = 0$. 这样一来, 我们得出结论:

$$H(f,\alpha) \cap P = \operatorname{conv}\{x_1, \cdots, x_k\},$$

从而 $F = H(f,\alpha) \cap P$ 是一个多面体.

最后集合 $\{x_1, \cdots, x_m\}$ 只有有穷多个子集, 并且 P 的每一个面对应于这种子集之一, 因此 P 只可能有有穷多个不同的面. ▮

下一个定理涉及多面体面的性质.

定理 1.7.3 设 $\{F_1, \cdots, F_k\}$ 是 $\mathbb{R}^n$ 中多面体 M 的一组面. 那么 $F_1 \cap \cdots \cap F_k$ 也是 M 的一个面.

证明 设 $F = F_1 \cap \cdots \cap F_k$. 如果 $F = \emptyset$, 则依据定义, F 总是面. 于是假定 $F \neq \emptyset$. 设 $x_0 \in F$, 注意每个 F_i 是 M 的真面, 从而存在 M 的承托超平面 $H(f_i, \alpha_i)$ 使得 $F_i = H(f_i, \alpha_i) \cap M$, $i = 1, \cdots, k$. 由于 $x_0 \in H(f_i, \alpha_i)$, $i = 1, \cdots, k$, 我们有

$$f_i(x_0) = \alpha_i, \quad i = 1, \cdots, k,$$
$$f_i(z) \geq \alpha_i, \quad \forall\, z \in M, i = 1, \cdots, k.$$

今定义 $\mathbb{R}^n$ 上的一个新的线性泛函 f,

$$f = f_1 + \cdots + f_k, \quad \text{并且 } \alpha = \alpha_1 + \cdots + \alpha_k.$$

现在我们证明 $H(f,\alpha)$ 是 M 的一个承托超平面，并且

$$H(f,\alpha) \cap M = F.$$

事实上，如果 $x \in F$，则 $x \in H(f_i,\alpha_i)$，$i = 1,\cdots,k$，从而 $f_i(x) = \alpha_i$. 因此，$f(x) = \sum_{i+1}^{k} f_i(x) = \sum_{i=1}^{k} \alpha_i = \alpha$，即 $x \in H(f,\alpha)$. 另一方面，若 $x \in M \backslash F$，则至少有一个指标 i 使得 $f_i(x) > \alpha_i$，从而 $f(x) = \sum_{i+1}^{k} f_i(x) > \sum_{i=1}^{k} \alpha_i = \alpha$，即 $x \notin H(f,\alpha) \cap M$. 因此 $F = H(f,\alpha) \cap M$.

最后，由于 $x_0 \in M \cap H(f,\alpha)$，并且 $f(x) \geq \alpha$, $\forall x \in M$，可知 $H(f,\alpha)$ 是 M 的一个承托超平面. ▮

在 1.6 节中我们看到，每一个闭凸集都是一族闭半空间的交. 对于多面体来说，我们将看到，只需有穷多个闭半空间就够了.

$\mathbb{R}^n$ 中的多面凸集是指有穷多个闭半空间之交.

定理 1.7.4 设 P 是 $\mathbb{R}^n$ 中的一个多面体，则 P 必是有界的多面凸集.

证明 设 $\{x_1,\cdots,x_m\}$ 为 P 的极小表现. 不失一般性，不妨假定 P 是 n 维多面体. 设 $F_1,\cdots,F_k$ 表示 P 的所有刻面. 对于每个 F_i，设 $H_i = H(f_i,\alpha_i)$ 为 P 的相应的承托超平面，使得

$$F_i = H_i \cap P, \quad P \subset H_i^+,$$

其中

$$H_i^+ = \{x \in \mathbb{R}^n \mid f_i(x) \geq \alpha_i\}.$$

于是 $P \subset H_1^+ \cap \cdots \cap H_k^+$. 今证 $P = H_1^+ \cap \cdots \cap H_k^+$. 我们用反证法来证明. 设不然，则有 $x \in \left(H_1^+ \cap \cdots \cap H_k^+\right) \backslash P$. 我们用 W 表示 $\{x_1, \cdots, x_m\}$ 中不超过 $n-1$ 个点组成的子集族，并令

$$M = \cap\{\mathrm{aff}(\{x\} \cap S) \mid S \in W\}.$$

M 是有穷多个维数至多为 $n-1$ 的仿射集之并. 由于 $\dim P = n$, 故 $\mathrm{int}\, P \subset M$ 不成立. 于是存在点 $y \in (\mathrm{int}\, P) \backslash M$. 这样我们有 $x \notin P$, $y \in \mathrm{int}\, P$, 所以在连接 x 和 y 的线段的内部有一点 $z \in \mathrm{bd}\, P$ (见图 1.7.2). 今证 z 属于 P 的某个刻面而不属于更低维数的面. 事实上，若 z 属于 P 的某个 j 维面， $0 \leq j \leq n-2$, 则根据 Caratheodory 定理 1.3.6, $z \in \mathrm{conv}\, S$, 其中 S 是 $\{x_1, \cdots, x_m\}$ 中不超过 $n-1$ 个点所组成的某个子集. 这表明 $z \in M$, 所以位于连接 x 和 z 的直线上的点 y 也应在 M 中，得出矛盾.

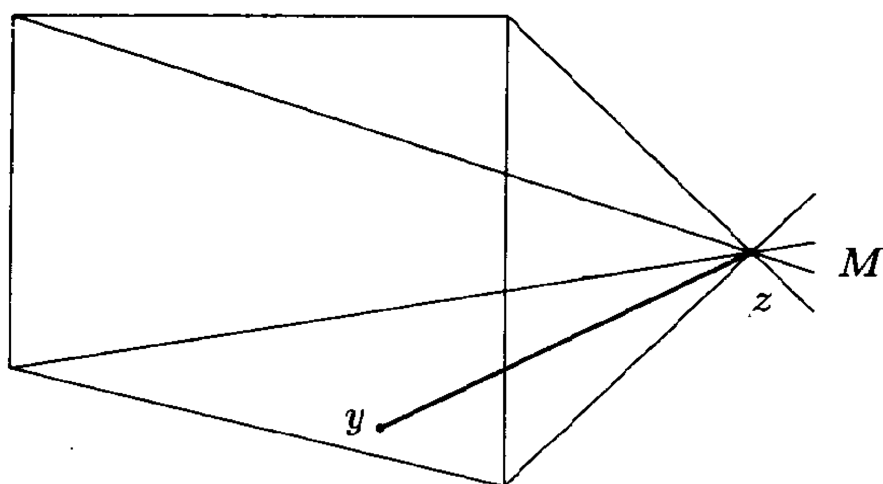


图 1.7.2

这样，我们得出结论：z 属于 P 的某个刻面，比如说，$z \in F_j$. 于是 $z \in H_j$, 并且由于 $y \in \operatorname{int} P \subset H_j^+$, x 不可能在 H_j^+ 中，这与我们一开始所作的假设 $x \in H_1^+ \cap \cdots \cap H_k^+$ 相矛盾. 因此必定有

$$H_1^+ \cap \cdots \cap H_k^+ \subset P.$$

这样我们证明了 $P = H_1^+ \cap \cdots \cap H_k^+$, 即 P 是一多面凸集，当然它是有界的. ∎

上述定理的逆也成立.

定理 1.7.5 $\mathbb{R}^n$ 中的每个有界多面凸集必定是一个多面体.

证明 设 P 是 $\mathbb{R}^n$ 中有界多面凸集. P 显然是紧集. 依据定理 1.6.5, 我们只需证明 P 只有有穷多个极点. 我们通过关于空间维数 n 的归纳法来证明这个结论.

对于 $n = 1$, P 或者是单点集，或者是一闭线段. 因此 P 至多有两个极点.

现在假设 $\mathbb{R}^{n-1}$ 中每个有界多面凸集均只有有穷多个极点. 考察 $\mathbb{R}^n$ 中的某个界多面凸集 P, 我们要证明 P 也只有有穷多个极点. 设 $H_1, \cdots, H_m$ 是 $\mathbb{R}^n$ 中的 m 个超平面，并且相应的闭半空间 H_j^+ 使得 $P = H_1^+ \cap \cdots \cap H_m^+$. 如果 x 是 P 的极点，则 $x \in \operatorname{bd} P$, 从而存在某个 $j, 1 \leq j \leq m$ 使得 $x \in H_j$(因为否则的话， $x \in \operatorname{int} P$). 但是我们知道 P 在 H_j 中的极点也是 $P \cap H_j$ 的极点 (见习题 1.6.3), 并且根据归纳假设， $P \cap H_j$ 只有有穷多个极点. 但构成 P 的超平面的个数是有限的，因此 P 只可能有有穷个极点. ∎

我们将在 3.5 节中进一步讨论无界多面凸集的一些基本

性质.

习 题 1.7

1.7.1 试证: $I\!R^n$ 中每个紧凸集 M 的顶点都是其极点; 并在 $I\!R^2$ 中举出一反例说明当 M 不是多面体时逆命题不再成立.

1.7.2 设 P 是 $I\!R^n$ 中多面体, 而 L 为 $I\!R^n$ 中的一仿射集. 试证 $P \cap L$ 是一个多面体.

1.7.3 设 $S \subset I\!R^n$ 为紧凸集, 并设 F_1 是 S 的一子集. 假定 $F_2 \subset F_1$ 并且 F_2 是 S 的一个面. 试证: F_2 是 F_1 的面.

1.7.4 设 $P_1, \cdots, P_m$ 是 $I\!R^n$ 中的多面体. 试证下列每一个集会也是一个多面体:

(1) $\operatorname{conv}(P_1 \cap \cdots \cap P_m)$;

(2) $P_1 \cap \cdots \cap P_m$;

(3) $P_1 + \cdots + P_m$.

1.7.5 设 P 是 $I\!R^n$ 的一个多面体, $\lambda > 0$. 试证: λP 也是一个多面体.

1.8 凸集的无界性

本节中我们讨论凸集的无界性. 直观上, $I\!R^n$ 中的无界凸集 M 在"无穷远"处有着比较简单的行为: 对于 M 中的任一点 x, M 必包含从 x 出发的某条半直线 $R(x; x+y) = \{x + \lambda y \mid \lambda \geq 0\}$, 这里 $y \in R^n, y \neq 0$. 注意, $I\!R^n$ 中的向量 y 可以看成是 $I\!R^n$ 中的点, 当 $y \neq 0$ 时也可以看成是"方向". 上面的半直线 $R(x; x+y)$ 中的 y 就是一个方向. 两条平行的半直线有相同的方向. 因此当两个方向相差某个正数倍时, 我们认为它们是同一个方向. 对于任一非空凸子集 $M \subset I\!R^n$ 以及 $y \in I\!R^n, y \neq 0$, 我们称 y 为 M 的一个回收方向 (direction of recession), 是指对于任意 $x \in M$, M

包含从 x 出发沿 y 方向的半直线，即

$$x+\lambda y \in M, \quad \forall \lambda \geq 0, \forall x \in M.$$

M 的所有回收方向加上零向量构成的集合叫做 M 的回收锥 (cone of recession), 记作 0^+M. 图 1.8.1 中的 y_1, y_2, y_3 都是回收方向.

从定义不难看出，有界凸集的回收锥仅由零向量构成.

对于任意非空凸子集 $M \subset I\!R^n$, 我们有

$$y \in 0^+M \Longleftrightarrow M+y \subset M.$$

事实上，若 $y \in 0^+M$, 则从 0^+M 的定义知 $x+y \in M$, $\forall x \in M$, 从而 $M+y \subset M$. 反之若 $M+y \subset M$, 则

$$M+2y=(M+y)+y \subset M+y \subset M.$$

由此用归纳法可知对于任意自然数 m 和 $x \in M$ 有 $x+my \in M$. 从而再根据 M 的凸性得知 $x+\lambda y \in M, \forall x \in M, \forall \lambda \geq 0$, 即 $y \in 0^+M$.

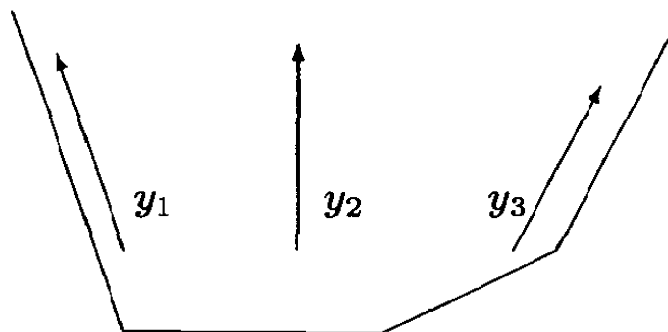


图 1.8.1　凸集的回收方向

此外，0^+M 确实是一个锥，因为当 $y \in 0^+M$ 时 $x+\lambda y \in M, \forall x \in M, \forall \lambda \geq 0$, 从而对于任意 $\mu \geq 0$ 也有 $x+\lambda(\mu y) \in M, \forall x \in M, \forall \lambda \geq 0$, 即 $\mu y \in 0^+M$. 由于 M

是凸集， 0^+M 还是一个凸锥. 事实上，若 $y_1, y_2 \in 0^+M$，$\lambda_1, \lambda_2 \geq 0$，则

$$\lambda_1 y_1 + \lambda_2 y_2 + M = \lambda_1 y_1 + (\lambda_2 y_2 + M) \subset \lambda_1 y_1 + M \subset M,$$

即 $\lambda_1 y_1 + \lambda_2 y_2 \in 0^+M$.

下面列举几个 $\mathbb{R}^2$ 中的回收锥的例子.

例 1.8.1 设 $M_1 = \{(\xi_1, \xi_2)^\tau \mid \xi_1 > 0, \xi_1\xi_2 \geq 1\}$，则

$$0^+M_1 = \{(\xi_1, \xi_2)^\tau \mid \xi_1 \geq 0, \xi_2 \geq 0\}.$$

例 1.8.2 设 $M_2 = \{(\xi_1, \xi_2)^\tau \mid \xi_2 \geq \xi_1^2\}$，则

$$0^+M_2 = \{(\xi_1, \xi_2)^\tau \mid \xi_1 = 0, \xi_2 \geq 0\}.$$

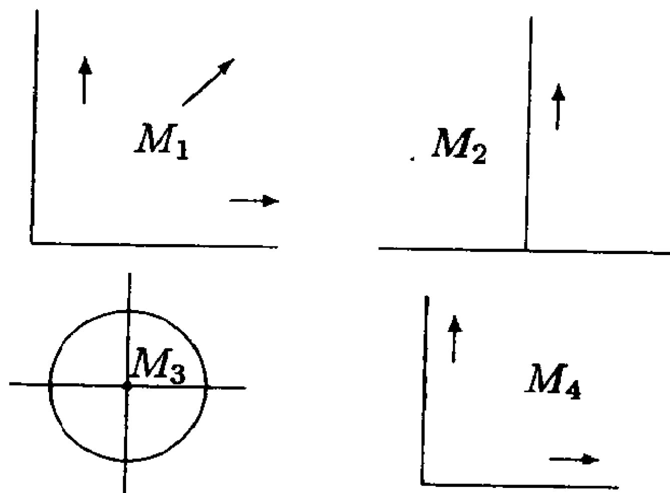


图 1.8.2 凸集的回收锥

例 1.8.3 设 $M_3 = \{(\xi_1, \xi_2)^\tau \mid \xi_1^2 + \xi_2^2 \leq 1\}$，则

$$0^+M_3 = \{(0, 0)^\tau\}.$$

例 1.8.4 设 $M_4 = \{(\xi_1, \xi_2)^\tau \mid \xi_1 > 0, \xi_2 > 0\} \cup \{(0, 0)^\tau\}$，则

$$0^+M_4 = 0^+M_1,$$

从而 $M_4 \subset 0^+M_4$.

例 1.8.5 设 $\{b_i \in \mathbb{R}^n \mid \beta_i \in \mathbb{R},\ i \in I\}$ 为给定的一族向量，令

$$M = \{x \in \mathbb{R}^n \mid \langle x, b_i \rangle \geq \beta_i, \forall\, i \in I\},$$

则

$$0^+M = \{x \in \mathbb{R}^n \mid \langle x, b_i \rangle \geq 0, \forall\, i \in I\}.$$

定理 1.8.1 设 M 是 $\mathbb{R}^n$ 中的一个非空闭凸集，那么

(1) 0^+M 是闭的，并且由一切形如 $\lambda_1 x_1, \lambda_2 x_2, \cdots$ 的序列极限构成，其中 $x_k \in M$, $\lambda_k \downarrow 0$ $(k \to \infty)$.

(2) 对于 $y \neq 0$, 如果存在一个 $x \in M$ 使得半直线 $\{x + \lambda y \mid \lambda \geq 0\}$ 包含在 M 中，则 $y \in 0^+M$; 此外，当 $x \in \operatorname{ri} M$ 时， $\{x + \lambda y \mid \lambda \geq 0\}$ 也包含在 $\operatorname{ri} M$ 内，从而 $y \in 0^+(\operatorname{ri} M)$.

证明 (1) 设 $\lambda_k \downarrow 0, x_k \in M$, $\lambda_k x_k \to y$ $(k \to \infty)$. 不妨假定 $0 < \lambda_k < 1$. 任取 $x \in M$, 从 M 的凸性知 $x + \lambda_k(x_k - x) \in M$, 并且 $\lambda_k(x_k - x) \to y$ $(k \to \infty)$. 于是从 M 的闭性得出 $x + y \in M, \forall x \in M$, 即 $y \in 0^+M$. 反之，设 $y \in 0^+M$, 并任取一元 $x \in M$. 对于一切自然数 k, 令 $x_k = x + ky$, $\lambda_k = 1/k$, 则 $x_k \in M$, $\lambda_k \downarrow 0$, 并且 $\lambda_k x_k = x/k + y \to y$ $(k \to \infty)$.

(2) 实际上，在 (1) 的证明中已包含了 (2) 的第一部分结论. 至于第二个结论可从定理 1.4.1 得出. ▮

推论 1.8.2 (1) 设 $M \subset \mathbb{R}^n$ 为任一非空凸集，则

$$0^+(\operatorname{ri} M) = 0^+(\operatorname{cl} M).$$

实际上对于任意给定的 $x \in \operatorname{ri} M$, 利用定理 1.4.1 有

$$y \in 0^+(\operatorname{cl} M) \Longrightarrow x + \lambda y \in \operatorname{ri} M, \quad \forall\, \lambda > 0.$$

(2) 若 $M \subset \mathbb{R}^n$ 为含原点的闭凸集，则

$$0^+M = \{y \mid y \in \lambda M, \forall \lambda > 0\} = \cap\{\lambda M \mid \lambda > 0\}.$$

(3) 设 $\{M_i \mid i \in I\}$ 为 $\mathbb{R}^n$ 的任意一族闭凸集，假定其交非空，那么

$$0^+(\cap \{M_i \mid i \in I\}) = \cap\{0^+M_i \mid i \in I\}.$$

(4) 设 $A: \mathbb{R}^n \to \mathbb{R}^m$ 为线性变换， $M \subset \mathbb{R}^m$ 为一闭凸集使得 $A^{-1}M \neq \emptyset$. 那么 $0^+(A^{-1}M) = A^{-1}(0^+M)$.

证明 留作练习. ∎

定理 1.8.3 设 $M \subset \mathbb{R}^n$ 为一非空闭凸集，则 M 为有界集，当且仅当 M 的回收锥 0^+M 仅由零向量构成.

证明 当 M 有界时，显然有 $0^+M = \{0\}$. 若 M 是无界集，则它含有一列向量 $x_1, x_2, \cdots$, 其范数 $\|x_k\| \to \infty$ $(k \to \infty)$. 若令 $\lambda_k = 1/\|x_k\|$, 则 $\lambda_k x_k$ 是单位向量，从而根据数学分析中关于有界集的 Weierstrass 定理， $\{\lambda_k x_k \mid k \geq 1\}$ 含一个子列收敛于 $\mathbb{R}^n$ 中某个单位向量 y. 于是依据定理 1.8.1, $y \in 0^+M$. ∎

推论 1.8.4 设 $M \subset \mathbb{R}^n$ 为一非空闭凸集，并设 $F \subset \mathbb{R}^n$ 是一仿射集，使得 $M \cap F$ 为非空有界. 那么对于每个与 M 平行的仿射集 F_1, $F_1 \cap M$ 也是有界的.

证明 从两个仿射集平行的定义可知 $0^+F = 0^+F_1$. 假定 $F_1 \cap M \neq \emptyset$, 那么

$$\begin{aligned} 0^+(F_1 \cap M) &= (0^+F_1) \cap 0^+M \\ &= (0^+F) \cap 0^+M = 0^+(F \cap M). \end{aligned}$$

于是从 $M \cap F$ 的有界性可得 $0^+(F_1 \cap M) = \{0\}$, 因此 $F_1 \cap M$ 有界. ∎

对于非空凸集 $M \subset I\!R^n$, 我们把集合 $(-0^+M) \cap 0^+M$ 称作 M 的直线状空间 (lineality space). 不难看出,

$$(-0^+M) \cap 0^+M = \{y \in I\!R^n \mid M + y = M\}.$$

M 的直线状空间是一个包含在凸锥 0^+M 中的最大的子空间, 其维数叫做 M 的直线状度 (lineality).

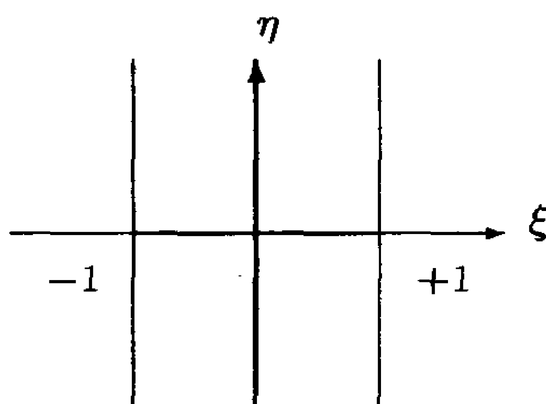


图 1.8.3　凸集的直线状空间

例如, 考虑平面上的条形区域

$$M = \{(\xi, \eta)^\tau \in I\!R^2 \mid |\xi| \le 1\},$$

则 M 的直线状空间是 η 轴, 故其直线状度等于 1 (见图 1.8.3). 一般说来, 对于非空凸集 $M \subset I\!R^n$, M 可以表示成如下直接和的形式:

$$M = L + (M \cap L^\perp),$$

其中 L 是 M 的直线状空间, $L^\perp$ 是 L 的正交补. 上式中 $M \cap L^\perp$ 的维数 (=M 的维数 $-M$ 的直线状度) 叫做 M 的秩, 这是 M 的非直线状的一种度量.

习　题　1.8

1.8.1 严格证明例 1.8.1–1.8.5 中的回收锥.

1.8.2 完成推论 1.8.2 的证明.

1.8.3 设 L 为 $\mathbb{R}^n$ 的一子空间，试证:$0^+L = L$.

1.8.4 设 $A:\mathbb{R}^n \to \mathbb{R}^m$ 为一线性变换， $y \in \mathbb{R}^m, \varepsilon > 0$. 令 $D_\varepsilon = \{x \in \mathbb{R}^n \mid \|y - Ax\| \le \varepsilon\}$. 试证:

$$0^+D_\varepsilon = N(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}.$$

1.9　一般凸集的面*

在 1.7 节中我们研究了凸多面体. 在那里把多面凸集定义成有穷多个闭半空间之交，并指出凸多面体就是有界的多面凸集. 本节我们给出一般凸集的面的定义和基本性质，这些结果将在 3.5 节中用来讨论无界多面凸集.

在 1.7 节中我们通过承托超平面定义了多面体的面. 考察 $\mathbb{R}^3$ 中某个闭圆环的凸包 M. 如果我们也把 M 的面定义成其承托超平面与 M 之交，则 M 的上、下圆盘的圆周上的点就不是 M 的“面”了 (见图 1.9.1). 因此，对于 $\mathbb{R}^n$ 中一般的凸集 M, 我们称 M 的一个凸子集 F 是 M 的一个面，是指它满足：对于 M 中任意闭线段 R, 如果 $F \cap \operatorname{ri} R \neq \emptyset$, 则必有 $R \subset F$. 由该定义可见， M 的零维面恰好是其极点. 设 H 是凸集 $M \subset \mathbb{R}^n$ 的闭包的一个 (非不足道的) 承托超平面，我们把 $H \cap M$ 叫做 M 的一个暴露面. 特别零维暴露面叫做暴露点. 这样， 1.7 节中定义的多面体的面和

顶点即是这里的暴露面和暴露点.

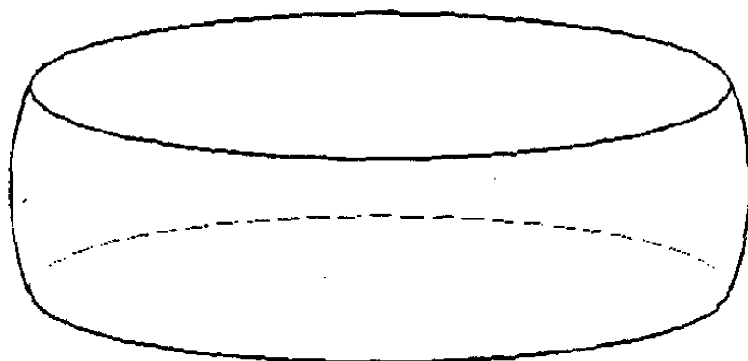


图 1.9.1

对于凸锥而言，零维面即极点的概念并没有多大的用处，因为其唯一的极点是零点. 实际上，凸锥的极射线概念更有意义. 凸锥 K 的极射线是指它既是 K 的一个面，又是从零点出发的一条半直线. 一般地，如果 R 是 $\mathbb{R}^n$ 中凸集 M 的一个半直线面 (即极射线), 则我们把 R 的方向叫做 M 的极方向 (M 在无穷远处的"极点"). 于是一个凸锥的极射线与该锥的极方向一一对应.

从凸集面的定义可以看出，若 F 是凸集 $M \subset \mathbb{R}^n$ 的一个面，而 S 是 F 的一个面，则 S 也是 M 的一个面. 特别地，凸集 M 的一个面的极点或极方向也是 M 本身的极点或极方向.

定理 1.9.1 设 $M \subset \mathbb{R}^n$ 为一非空凸集， F 是 M 的一个面. 如果 D 是 M 的一个凸子集使得 $F \cap \operatorname{ri} D \neq \emptyset$, 那么 $D \subset F$.

证明 设 $z \in F \cap \operatorname{ri} D$. 如果 x 是 D 中任意点且 $x \neq z$, 则存在一点 $y \in D$, 使得 z 位于连接 x 和 y 的线段的相对内

部. 由于 F 是 M 的一个面，故必有 $x,y\in F$, 从而 $D\subset F$. ▮

推论 1.9.2 设 $M\subset \mathbb{R}^n$ 为一非空凸集，那么，

(1) 对于 M 的非空面 F, 有 $F=M\cap \mathrm{cl}\,F$. 特别当 M 是闭凸集时， F 是闭的；

(2) 如果 F_1 和 F_2 是 M 的两个面，满足 $\mathrm{ri}\,F_1\cap \mathrm{ri}\,F_2\neq\emptyset$, 则 $F_1=F_2$;

(3) 如果 F 是 M 的一个真面，则 F 完全包含在 M 的相对边界内，即 $F\subset \mathrm{bd_r}\,M$, 从而 $\dim F<\dim M$.

证明 (1) 在定理 1.9.1 中取 $D=M\cap \mathrm{cl}\,F$.

(2) 从 $F_1\cap \mathrm{ri}\,F_2\neq\emptyset$ 推出 $F_2\subset F_1$, 类似地从 $F_2\cap \mathrm{ri}\,F_1\neq\emptyset$ 可推出 $F_1\subset F_2$.

(3) 若 $F\cap \mathrm{ri}\,M\neq\emptyset$, 则必 $M\subset F$, 但 F 是 M 的真凸子集，这不可能. ▮

设 $\mathcal{F}(M)$ 表示 $\mathbb{R}^n$ 中给定凸子集 M 的所有面组成的集族. 于是 $\mathcal{F}(M)$ 按集合包含关系 $\subseteq$ 构成一个半序集. $\mathcal{F}(M)$ 有最大元 (即 M) 和最小元 (即 $\emptyset$). 此外 $\mathcal{F}(M)$ 的任意一组面之交仍然是 M 的面. 因此 $\mathcal{F}(M)$ 的任意一组面有最大下界 (按包含半序关系), 同样也有最小上界. 此外，任意一列严格递减的面只能有有穷多个，这是因为根据推论 1.9.2, 这一列面的维数是严格递减的.

定理 1.9.3 设 M 是 $\mathbb{R}^n$ 中非空凸集，并设是 $\mathcal{G}(M)$ 为 M 的所有非空面的相对内部构成的集族. 那么 $\mathcal{G}(M)$ 是 M 的一个划分，即 $\mathcal{G}(M)$ 中的集合互不相交，并且其并等于 M. 此外， M 的每一个相对开凸子集必包含在 M 的某个面的相对内部中； $\mathcal{G}(M)$ 中的每个集合 G 是 M 的极大相对开凸集，即不存在 M 的相对开凸集 O 使得 O 真包含

G.

证明 依据推论 1.9.2, M 的不同面的相对内部互不相交. 任意取 M 的一个非空开凸子集 D (例如, D 可以是单点集), 设 F 是 M 的包含 D 的最小的面 (即 M 的所有包含 D 的面之交). 如果 D 包含在 F 的相对边界 $\mathrm{bd_r}\, F$ 中, 则存在 F 的一个承托超平面 H 使得 H 包含 D 但不包含整个 F. 于是 D 应该位于 F 的暴露面 $F_1 = F \cap H$ 中, 这就是说, M 是一个比 F 小的真面, 并且显然 $F_1 \supset D$, 从而与 F 是 M 的包含 D 的最小面的假设矛盾. 因此, D 不可能完全包含在 $\mathrm{bd_r}\, F$ 中, 从而 $D \cap \mathrm{ri}\, F \neq \emptyset$. 于是注意到,

$$\mathrm{ri}\, D \cap \mathrm{ri}\, F = \mathrm{ri}\, D \cap \mathrm{ri}\,(\mathrm{cl}\, F) = \mathrm{ri}\,(D \cap \mathrm{cl}\, F) = \mathrm{ri}\, F,$$

可知 $D = \mathrm{ri}\, D \subset \mathrm{ri}\, F$. 这就是说, D 包含在 $\mathcal{G}(M)$ 的某个成员中. 注意 $\mathcal{G}(M)$ 的不同集合必不相交, 因此 $\mathcal{G}(M)$ 中的集合必定是 M 的极大开凸子集, 并且其并等于 M. ∎

设 S_0 是 $\mathbb{R}^n$ 中的点集, 而 S_1 是 $\mathbb{R}^n$ 中的一些方向组成的集合. 我们定义集合 $S = S_0 \cup S_1$ 的凸包 $\mathrm{conv}\, S$ 为满足如下条件的最小的凸集 M: $M \supset S_0$ 并且 $S_1 \subset O^+M$. 显然这样的最小凸集 M 是存在的. 事实上,

$$M = \mathrm{conv}\, S_0 + \mathrm{cone}\, S_1,$$

其中 $\mathrm{conv}\, S_0$ 是点集 S_0 在通常意下的凸包, 而 $\mathrm{cone}\, S_1$ 则是由 S_1 作为点集所生成的凸锥. 因此向量 $z \in \mathrm{conv}\, S$ 当且仅当

$$z = \lambda_1 x_1 + \cdots + \lambda_k x_k + \mu_1 y_1 + \cdots + \mu_m y_m,$$

其中 $x_i \in S_0, y_j \in S_1$, $\lambda_i \geq 0, \mu_j \geq 0$, $1 \leq i \leq k, 1 \leq j \leq m$, $\lambda_1 + \cdots + \lambda_k = 1$. 我们把这样的向量 z 叫做 S 中 $k+m$ 个点和方向的凸组合. 这样的凸组合恰好对应于

$\mathbb{R}^{n+1}$ 中的位于超平面 $H=\{(x,1)\in\mathbb{R}^{n+1}\mid x\in\mathbb{R}^n\}$ 的非负线性组合

$$\sum_{i=1}^{k}\lambda_k(x_i,1)+\sum_{j=1}^{m}\mu_j(y_j,0).$$

于是 $\operatorname{conv}S$ 也可以定义成 $\mathbb{R}^{n+1}$ 中超平面 H 与由 S' 生成的凸锥之交，其中 $S'=\{(x,1)\mid x\in S_0\}\cup\{(y,0)\mid y\in S_1\}$.

如果 $M=\operatorname{conv}S$, 则 M 的面与 S 的某些子集之间有着一一对应的关系. 我们有如下定理:

定理 1.9.4 设 $M=\operatorname{conv}S$, 其中 S 是由 $\mathbb{R}^n$ 中的点集 S_0 和方向集 S_1 组成的集合，并设 F 是 M 的一个非空面. 那么 $F=\operatorname{conv}S'$, S' 为由 S 中属于 F 的点以及 S 中属于 O^+F 的回收方向所组成的集合，即 $S'=S_0'\cup S_1'$, 其中 $S_0'=F\cap S_0$, 而 $S_1'=(O^+F)\cap S_1$.

证明 根据定义，我们有 $F\supset\operatorname{conv}S'$. 另一方面，设 $z\in F$, 我们证明 $z\in\operatorname{conv}S'$. 事实上，由于 $z\in\operatorname{conv}S$, 存在点 $x_1,\cdots,x_k\in S_0$ 和方向 $y_1,\cdots,y_m\in S_1$, 使得

$$z=\lambda_1x_1+\cdots+\lambda_kx_k+\mu_1y_1+\cdots+\mu_my_m,$$

其中 $\lambda_i,\mu_j>0,\forall 1\le i\le k,1\le j\le m$, $\lambda_1+\cdots+\lambda_k=1$. 现在设 S'' 由点 $x_1,\cdots,x_k$ 和方向 $y_1,\cdots,y_m$ 组成，并记 $D=\operatorname{conv}S''$. 于是，由于上述表达式中诸系数全是正的，故依据定理 1.4.11, $z\in\operatorname{ri}D$. 由此可见 $\operatorname{ri}D$ 与 F 相交. 根据定理 1.8.1, $D\subset F$, 从而 $x_1,\cdots,x_k\in F$, 并且当 $m\ge1$ 时 F 包含以 $y_1,\cdots,y_m$ 为方向的一些半直线. 从而 $y_1,\cdots,y_m$ 是 $\operatorname{cl}F$ 的回收方向，同时也是 M 的回收方向. 但由推论 1.9.2 知 $F=M\cap\operatorname{cl}F$, 因此 $y_1,\cdots,y_m$ 实际上也是 F 的回收方向. 这样，我们证明了 $S''\subset S'$, 并且 $z\in\operatorname{conv}S'$. ▌

推论 1.9.5 设 S 是 $\mathbb{R}^n$ 中一些点和方向组成的集合，记 $M = \operatorname{conv} S$. 那么 M 的每个极点都是 S 中的点. 如果 M 中没有一条半直线含有 S 的无界点集 (特别当 S 中所有点有界时就是这种情形), 则 M 的每个极方向必是 S 中的方向.

证明 在定理 1.9.4 中取 F 为由单点或半直线构成的 M 的一个面. ∎

当一个凸集 $M \subset \mathbb{R}^n$ 的直线状度不等于零时，显然 M 没有极点或极方向. 然而在这种情况下， M 可以分解成 $M = M_0 + L$, 其中 L 为 M 的直线状子空间，而 $M_0 = M \cap L^{\perp}$ 为直线状度等于零的凸子集. M 的面 F 显然与 M_0 的面 F_0 之间有一一对应：$F = F_0 + L$, $F_0 = F \cap L^{\perp}$. 因此在研究凸集的面的时候，只要研究具有直线状度为零的凸集的面就清楚了.

在 1.6 节中我们已指出任意有界闭凸集是其极点集的凸包，从而更是其相对边界的凸包. 但是对于仿射集或闭半仿射集 (可类似于闭半空间定义), 这个结论显然是不成立的. 但是我们有

定理 1.9.6 设 M 是 $\mathbb{R}^n$ 中的一个闭凸集. 如果 M 既不是一仿射集也不是一闭半仿射集，那么 M 的每一相对内点必包含在连接 M 的两个相对边界点的某直线段上.

证明 设 $D = \operatorname{bd_r} M$. 由于 M 不是仿射集， D 必非空. D 不可能是凸集，因为否则的话，根据定理 1.5.4, 存在一超平面 H 把 D 和 $\operatorname{ri} M$ 真分离，并且 $D \subset H$. 设 H^+ 是包含 $\operatorname{ri} M$ 但与 D 不相交的相应的开半空间. 由于 M 不是闭半仿射集，必存在点 $x \in H^+ \cap \operatorname{aff} M$ 使得 $x \notin \operatorname{ri} M$. 连接 x 和 $\operatorname{ri} M$ 中任意点的线段与 M 之交必是一端在 D 中

的一线段. 但这是不可能的，因为 H^+ 与 D 不相交. 因此 D 不是凸集，并且必存在 D 中两个不同点 x 和 y, 使其连线含有 $\mathrm{ri}\, M$ 中的点. 设 R 是通过 x 和 y 的直线，则 R 与 M 的交必定是连接 x 和 y 的线段，因为否则的话，根据定理 1.4.1, x 或 y 必定属于 $\mathrm{ri}\, M$ 了. 根据推论 1.8.4, 任何平行于 R 的直线与 M 的交也是闭有界的. 这样一来，给定任意点 $y \in \mathrm{ri}\, M$, 通过 y 并平行于 R 的直线与 M 之交是一线段，其两个端点在 D 中. ∎

下一个定理是定理 1.6.5 的推广.

定理 1.9.7 设 M 是 $\mathbb{R}^n$ 中不含任何直线的闭凸集, 并设 S 是其所有极点和极方向组成的集合. 那么 $M = \mathrm{conv}\, S$.

证明 如果 $\dim M \le 1$, 则显然定理成立, 因为这时 M 或者是 $\emptyset$, 或者是一单点集，或者是一闭线段，或者是一闭半直线. 我们用归纳法来证明. 设定理对于维数 $< m$ $(m > 1)$ 的闭凸集成立. 假定 $\dim M = m$. 注意 S 中的点属于 M, 而 S 中的极方向是 M 的回收方向，因此显然有 $M \supset \mathrm{conv}\, S$. 由于 M 不含直线并且本身又不是半直线, 故依据定理 1.9.6, M 是其相对边界 $\mathrm{bd_r}\, M$ 的凸包. 因此为证 $M \subset \mathrm{conv}\, S$, 只需证明 $\mathrm{bd_r}\, M \subset \mathrm{conv}\, S$. 但根据定理 1.9.3, 对于任意 $x \in \mathrm{bd_r}\, M$, 必有 M 的某个真面 F 使得 $x \in \mathrm{ri}\, F$. 根据推论 1.9.2, F 是闭的，并且 $\dim F < \dim M$. 又根据归纳假设, 定理对于闭凸集 F 成立, 从而 $x \in \mathrm{conv}\, S_F$, 其中 S_F 是 F 的所有极点和极方向组成的集合. 但我们知道 $S_F \subset S$, 所以 $x \in \mathrm{conv}\, S$. ∎

推论 1.9.8 $\mathbb{R}^n$ 中不含任何直线的非空闭凸集至少有一个极点.

证明 如果定理 1.9.7 中的集合 S 仅有方向组成，则

$M = \operatorname{conv} S$ 是一个含原点的凸锥，从而原点又是 M 的极点了. ∎

定理 1.9.9 (Strasjewicz定理) 设 M 是 $\mathbb{R}^n$ 中的闭凸集，那么 M 的暴露点集是其极点集的一个稠子集.

证明 首先注意对于任意 $\alpha > 0$, $x \in B(0,\alpha)$ 是 M 的极点或暴露点，等同于 x 是 $M \cap B(0,\alpha)$ 的极点或暴露点. 因此，只需在 M 为有界且非空的情形下来证明定理就够了. 设 S 为 M 的暴露点集，于是，当然 S 包含在 M 的极点集 $E(M)$ 中，并且 $\operatorname{cl} S \subset M$. 我们要证明 M 的每个极点属于 $\operatorname{cl} S$, 即 $E(M) \subset \operatorname{cl} S$. 如不然，则存在 $x \in E(M) \backslash \operatorname{cl} S$, 于是 $x \notin M_0 \triangleq \operatorname{conv}(\operatorname{cl} S)$. 但 M_0 是闭的，故存在一闭半空间 G 包含 M_0 但不含 x. 下面我们来构造出 M 的一个不在 G 中的暴露点，从而导致矛盾.

设 b 表示 G 的"外"法向单位向量，并设 $\varepsilon > 0$ 是使得 $x - \varepsilon b \in G$ 的最小正数. 对于某个 $\lambda > \varepsilon$, 记 $y = x - \lambda b$. 显然 $B(y,\lambda)$ 的边界包含 x. 然而 G 中不属于 $B(y,\lambda)$ 的点到 x 的距离至少是 $(2\lambda\varepsilon)^{1/2}$. 令 $r = \sup\{\|z - x\| \mid z \in M \cap G\}$, 并选择 λ 充分大使得还满足 $(2\lambda\varepsilon)^{1/2} > r$. 于是若记 $D = \{z \in M \mid \|z - y\| \geq \lambda\}$, 则 $x \in D$, 并且 $M \cap G$ 与 D 不相交. 由于 M 是有界闭凸集，故存在 $p \in M$ 使得 $\|y - p\| = \sup\{\|z - y\| \mid z \in M\}$. 显然 $p \notin G$. 这样，若记 $\mu = \|y - p\|$, 则闭球 $\overline{B}(y,\mu)$ 在其边界点 p 处的 (唯一的) 承托超平面与 $\overline{B}(y,\mu)$ 之交只含点 p. 由于 $p \in M \subset \overline{B}(y,\mu)$, 可见 p 是 M 的暴露点. ∎

定理 1.9.10 设 M 是 $\mathbb{R}^n$ 的不含任何直线的闭凸集，并设 S 是 M 的所有暴露点和暴露方向的集合，那么 $M = \operatorname{cl}(\operatorname{conv} S)$.

证明 为简单起见，不妨假定 $\dim M = n$, 并且 $n \geq 2$ (当 $n < 2$ 时定理显然成立). 依据定理 1.9.7, 有 $M \supset \mathrm{cl}\,(\mathrm{conv}\, S)$. 又根据定理 1.9.7, $\mathrm{cl}\,(\mathrm{conv}\, S)$ 是包含 M 的所有极点的闭凸集，因此根据推论 1.9.8, $\mathrm{cl}\,(\mathrm{conv}\, S)$ 非空. 假定 $\mathrm{cl}\,(\mathrm{conv}\, S) \neq M$, 由此我们将导出矛盾. 事实上，根据凸集分离定理, 存在一个超平面 H 与 M 相交但与 $\mathrm{cl}\,(\mathrm{conv}\, S)$ 不相交. 于是从推论 1.9.8 可知凸集 $M \cap H$ 至少存在一极点，因此也至少有一暴露点 x (定理 1.9.9). 根据暴露点的定义，在 H 中存在一 $(n-2)$ 维仿射集 G 使之与 $M \cap H$ 仅在 x 处相交. 特别是 G 不会与 M 的内部 $\mathrm{int}\, M$ 相交, 从而根据定理 1.5.4, 可以把 G 扩张成一个超平面 H' 使之仍与 $\mathrm{int}\, M$ 不相交. H' 是 M 的一个承托超平面，并且 $M' = M \cap H'$ 是 M 的一个暴露面. M' 的极点或暴露点总是 M 的极点，从而必属于 $\mathrm{cl}\,(\mathrm{conv}\, S)$ 但不属于 H. 超平面 H 与 M' 仅仅相交于 x 处. 由于 x 不可能是 M' 的暴露点, 故必有 $H \cap \mathrm{ri}\, M' = \{x\}$, 因此 $\dim M' = 1$. 依据假设， M' 不可能是直线. 又 M' 也不可能是两点间的直线段，因为否则的话，这两点作为 M' 的极点将属于 S, 与 $x \notin \mathrm{conv}\, S$ 矛盾. 这样，唯一的可能是 M' 为端点在 S 中的闭半直线. 根据定义， M' 的方向是暴露方向，从而 M' 的方向属于 S. 这样我们得出结论: $M' \subset \mathrm{conv}\, S$, 与 $x \notin \mathrm{conv}\, S$ 矛盾. ∎

推论 1.9.11 设 K 是 $\mathbb{R}^n$ 中的闭凸锥，并且 K 不至含原点，但不含任何直线. 设 W 表示 K 的所有暴露射线之并，那么 $K = \mathrm{cl}\,(\mathrm{conv}\, W)$.

习 题 1.9

1.9.1 设 M 为 $\mathbb{R}^n$ 中的闭凸集，试证: $y \in \mathbb{R}^n$ 是 M 的极方

向当且仅当 y 不可能表示成 M 的两个不同的回收方向的正线性组合.

1.9.2 设 M 为 $\mathbb{R}^n$ 中的闭凸集，试证：M 的暴露点必是其极点.

1.9.3 设 M 为 $\mathbb{R}^n$ 中的闭凸集，$x \in M$, 试证：x 是 M 的极点，当且仅当 x 是 $M \cap B(x,\delta)$ $(\delta > 0)$ 的极点；而 x 是 M 的暴露点当且仅当 x 是 $M \cap B(x,\delta)$ 的暴露点.

1.9.4 设 M 为 $\mathbb{R}^n$ 中的紧子集，并设 H 为 $\mathbb{R}^n$ 的某个开半空间. 试证：若 $H \cap M \neq \emptyset$, 则在 H 中至少有 M 的一个暴露点.

1.9.5 举例说明非暴露的一维面是存在的.

第二章　凸函数

凸函数是凸分析中重要的研究对象之一. 它研究的内容非常丰富，研究的结果已在许多领域得到广泛应用. 其中一个重要的应用是关于研究优化问题，包括线性规划和凸规划问题等. 当然，这些理论的完整的叙述已超出本书的研究范围，有兴趣的读者可参考有关的专著. 我们将在第五章讨论关于规划问题的一些应用.

在讨论 $\mathbb{R}^n$ 上的凸函数之前，我们先介绍单变量凸函数，以便读者对凸函数有一个直观的概念. 我们将指出如何通过上图概念把凸函数与凸集联系起来.

2.1　单变量凸函数

本节中我们始终假定 $-\infty \leq a < b \leq +\infty$，于是 (a,b) 表示某个有界或无界区间.

我们称函数 $f:(a,b) \to \mathbb{R}$ 为凸函数，是指它满足

$$\begin{cases} f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2), \\ \qquad \forall\, x_1, x_2 \in (a,b),\ 0 \leq \lambda \leq 1. \end{cases} \tag{2.1.1}$$

几何上，f 是凸函数意味着 f 的图象是下凸的 (见图 2.1.1). 如果 (2.1.1) 中的不等号对一切 $x \in (a,b)$ 严格成立，则称 f 是 (a,b) 上的严格凸函数. 当 f 是 (a,b) 上的凸函数时，有时也称 $-f$ 是 (a,b) 上的凹函数.

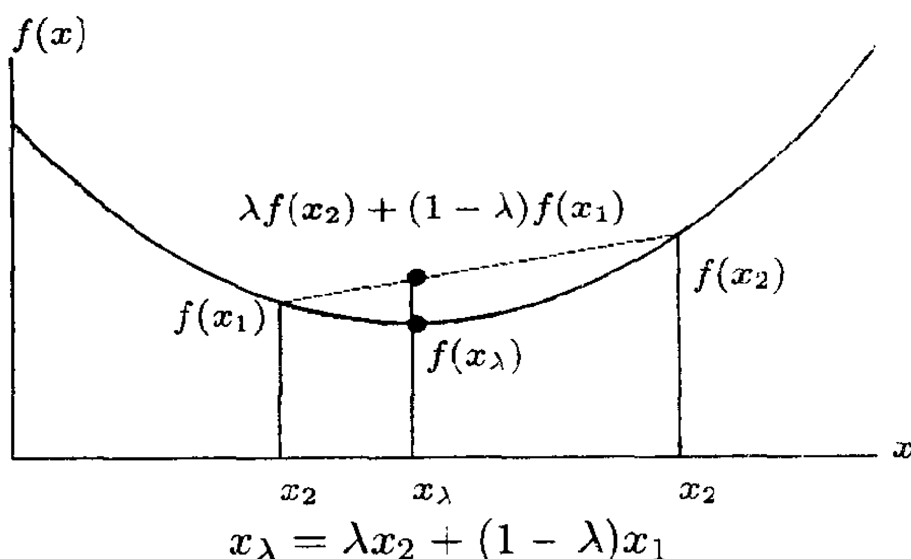


图 2.1.1 单变量凸函数

定理 2.1.1 (Jensen不等式) 设 $f:(a,b)\to\mathbb{R}$, 那么为了 f 是凸函数必须且只须对于任意自然数m, 成立下列不等式:

$$\begin{cases} f\left(\sum_{k=1}^{m}\lambda_k x_k\right)\le\sum_{k=1}^{m}\lambda_k f(x_k), \\ \forall\, x_k\in(a,b),\forall\,\lambda_k\ge 0,\ \lambda_1+\cdots+\lambda_m=1. \end{cases} \tag{2.1.2}$$

证明 充分性不待证. 我们只需证必要性. 设 f 是 (a,b) 上的凸函数，我们用数学归纳法证明式 (2.1.2) 成立. 对于 $m=2$, 式 (2.1.2) 正好是凸函数的定义. 假定式 (2.1.2) 对于 $m=k\ge 2$ 成立，现在证明 $m=k+1$ 时式 (2.1.2) 也成立. 为此，任取非负数 $\lambda_1,\cdots,\lambda_{k+1}$ 满足 $\lambda_1+\cdots+\lambda_k=1$, 并取 $x_1,\cdots,x_{k+1}\in(a,b)$. 不妨假定 $0<\lambda_{k+1}<1$. 令 $\mu=\lambda_1+\cdots+\lambda_k$, $\lambda'_k=\lambda_k/\mu, k=1,\cdots,k$. 由于 $\lambda'_1+\cdots+\lambda'_k=1$, 故由归纳假设可得

$$f(\lambda'_1x_1+\cdots+\lambda'_kx_k)\le\lambda'_1f(x_1)+\cdots+\lambda'_kf(x_k).$$

考虑到 $\mu+\lambda_{k+1}=1$, 再由 f 的凸性的定义即得

$$
\begin{aligned}
& f(\lambda_1 x_1+\cdots+\lambda_{k+1}x_{k+1}) \\
= & f(\mu(\lambda_1' x_1+\cdots+\lambda_k' x_k+\lambda_{k+1}x_{k+1}) \\
\le & \mu f(\mu(\lambda_1' x_1+\cdots+\lambda_k' x_k)+\lambda_{k+1}f(x_{k+1}) \\
\le & \lambda_1 f(x_1)+\cdots+\lambda_k f(x_k)+\lambda_{k+1}x_{k+1}.
\end{aligned}
$$

从而式 (2.1.2) 对于 $m=k+1$ 成立. 于是根据数学归纳法, 式 (2.1.2) 对于一切 m 成立. ∎

定理 2.1.2 给定函数 $f:(a,b)\to \mathbb{R}$, 对于任意固定的 $y\in(a,b)$, 定义函数

$$
\varphi_y(x)=\frac{f(x)-f(y)}{x-y},\quad x\neq y, x\in(a,b).
$$

那么为了 f 是 (a,b) 上的凸函数, 必须且只须对于任意固定的 $y\in(a,b)$, $\varphi_y(x)$ 作为 x 的函数在 (a,b) 上是不减的 (除了 $x=y$).

证明 函数 $\varphi_y(x)$ 不减意味着对于任意 $x_1,x_2\in(a,b)\backslash\{y\}$, $x_1<x_2$, 有

$$
\frac{f(x_1)-f(y)}{x_1-y}\le\frac{f(x_2)-f(y)}{x_2-y}. \tag{2.1.3}
$$

为了确定起见, 这里我们假定 $x_2>x_1>y$, 至于情况 $x_2>y>x_1$ 和 $y>x_2>x_1$ 的讨论是完全类似的. 于是存在 $\lambda\in\mathbb{R}$, $0<\lambda<1$ 使得 $x_1=(1-\lambda)y+\lambda x_2$. 由此式 (2.1.3) 等价于

$$
\frac{f(x_1)-f(y)}{\lambda(x_2-y)}\le\frac{f(x_2)-f(y)}{x_2-y},
$$

即

$$f(x_1)=f((1-\lambda)y+\lambda x_2)\le(1-\lambda)f(y)+\lambda f(x_2).$$

由于 y 的任意性，这正表明 f 是 (a,b) 上的凸函数. ▌

定理 2.1.3 设 $f:(a,b)\to\mathbb{R}$ 为凸函数，那么 f 在 (a,b) 上处处左、右可微，从而是 (a,b) 中的连续函数. 此外，其左、右导数 f'_- 和 f'_+ 满足

$$\begin{aligned}&f'_-(x_1)\le f'_+(x_1)\le\frac{f(x_2)-f(x_1)}{x_2-x_1}\le f'_-(x_2)\\&\le f'_+(x_2),\ \forall\, x_1,x_2\in(a,b),\ x_1<x_2.\end{aligned}\tag{2.1.4}$$

证明 事实上，f 的左、右导数的存在性由定理 2.1.2 中函数 $\varphi_y(x)$ 的不减性立即得知. 当 $x>x_2>y$ 时，从定理 2.1.2 可知

$$\frac{f(x)-f(x_2)}{x-x_2}\le\frac{f(y)-f(x_2)}{y-x_2}.$$

于是

$$f'_+(x_2)=\lim_{x\downarrow x_2}\frac{f(x)-f(x_2)}{x-x_2}\ge\lim_{y\uparrow x_2}\ge f_-(x_2).$$

同理，当 $x_2>x>x_1$ 时，有

$$\frac{f(x)-f(x_2)}{x-x_2}\ge\frac{f(x)-f(x_1)}{x-x_1}.$$

因此

$$\begin{aligned}f'_-(x_2)&=\lim_{x\uparrow x_2}\frac{f(x)-f(x_2)}{x-x_2}\ge\frac{f(x_2)-f(x_1)}{x_2-x_1}\\&\ge\lim_{x\downarrow x_1}\frac{f(x)-f(x_1)}{x-x_1}=f'_+(x_1).\end{aligned}$$

综合上述几个不等式即得 (2.1.4). ▌

推论 2.1.4 设 $f:(a,b)\to\mathbb{R}$ 为凸函数，那么,

(1) f 的左、右导数 f'_- 和 f'_+ 均是 (a,b) 上的不减函数；

(2) f 在 (a,b) 上的任一闭子区间 $[c,d]$ 上是 Lipschitz 函数，即存在常数 $K>0$ 使得

$$|f(x_1)-f(x_2)| \le K|x_1-x_2|, \quad \forall\, x_1,x_2 \in [c,d]. \tag{2.1.5}$$

证明 (1) 是式 (2.1.4) 的直接结论. 为证 (2), 不妨设 $x_1,x_2 \in [c,d]$, $x_1 < x_2$. 于是由式 (2.1.4) 和 (1) 可知,

$$f'_-(c) \le f'_-(x_1) \le \frac{f(x_2)-f(x_1)}{x_2-x_1} \le f'_-(x_2) \le f'_-(d).$$

从而只要取 $K=\max\{|f'_-(c)|,|f'_-(d)|\}$, 就有式 (2.1.5) 成立. ∎

推论 2.1.5 设 $f:(a,b)\to \mathbb{R}$ 为凸函数，那么 f 在 (a,b) 上除了至多可数个点外处处可微.

证明 f 在 (a,b) 的内点 x_0 处不可微意味着

$$f'_-(x_0) < f'_+(x_0).$$

从式 (2.1.4) 可知，对于 f 的任意两个不可微的点 $x_1,x_2 \in (a,b)$, 两个区间 $(f'_-(x_1),f'_+(x_1))$ 和 $(f'_-(x_2),f'_+(x_2))$ 不相交，从而这样的区间至多只有可数多个. ∎

定理 2.1.3 的逆命题也成立. 为此我们先证明一个引理.

引理 2.1.6 设连续函数 $f:(a,b)\to \mathbb{R}$ 处处右可微，那么对于任意 $x_1,x_2\in(a,b), x_1<x_2$, 有

$$\inf_{x_1<x<x_2} f'_+(x) \le \frac{f(x_2)-f(x_1)}{x_2-x_1} \le \sup_{x_1<x<x_2} f'_+(x). \tag{2.1.6}$$

证明 我们先证明式 (2.1.6) 的第二个不等号成立. 令

$$M \triangleq \sup_{x_1<x<x_2} f'_+(x), \quad m \triangleq \inf_{x_1<x<x_2} f'_+(x).$$

如果 $M=+\infty$, 则式 (2.1.6) 第二个不等号当然成立. 因此只需考虑 M 为有穷的情形. 令

$$g(x) = Mx - f(x),$$

显然 g 右可微，并且

$$g'_+(x) \geq 0, \quad \forall\, x \in (x_1, x_2).$$

今证 g 是 (x_1, x_2) 中的不减函数，即

$$g(z_2) \geq g(z_1), \quad \forall\, z_1, z_2 \in (x_1, x_2),\ z_2 > z_1. \qquad (2.1.7)$$

事实上， $\forall\, x \in (x_1, x_2), \forall\, \varepsilon > 0$, 存在 $\delta_x > 0$ 使得

$$g(x+h) - g(x) \leq -\varepsilon h, \quad \forall\, h \in [0, \delta_x].$$

现在设

$$z \triangleq \sup\{x \in [z_1, z_2] \mid g(x) - g(z_1) \geq -\varepsilon(x - z_1)\}.$$

我们来证明 $x = z_2$. 如果不然，则有 $z < z_2$, 从而由 z 的定义和 g 的连续性可知

$$g(z) - g(z_1) \leq -\varepsilon(z - z_1)..$$

但对于 z, 又有 $\delta_z > 0$ 使得

$$g(z+h) - g(z) \geq -\varepsilon h, \quad \forall \in [0, \delta_z],$$

这显然与 z 的定义相抵触. 于是由 $\varepsilon > 0$ 的任意性知式 (2.1.7) 成立. 根据 g 的定义，式 (2.1.7) 也可写成

$$Mz_2 - f(z_2) \geq Mz_1 - f(z_1), \quad \forall\, z_1, z_2 \in (x_1, x_2), z_1 < z_2.$$

由此利用 f 的连续性得到

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} \geq M = \sup_{x_1<x<x_2} f'_+(x)$$

为了证明 (2.1.6) 的第一个不等号，令 $p(x) = f(x) - mx$，并作与上述类似的推理. ∎

定理 2.1.7 为了 $f:(a,b) \to I\!R$ 是凸函数，必须且只须 f 在 (a,b) 上处处左、右可导，并且其左、右导数 f'_- 和 f'_+ 满足式 (2.1.4).

证明 只需证明充分性. 事实上，在式 (2.1.4) 成立的条件下，从引理 2.1.6 可知

$$\begin{cases} \dfrac{f(x_1)-f(x)}{x_1-x} \leq \sup\limits_{x_1<y<x} f'_+(y) \\ \leq \inf\limits_{x<y<x_2} f'_+(y) \leq \dfrac{f(x_2)-f(x)}{x_2-x}, \\ \forall x_1, x_2 \in (a,b)\ x_1 < x_2,\ \forall x \in (x_1,x_2). \end{cases}$$

于是根据定理 2.1.2, f 是 (a,b) 上的凸函数. ∎

推论 2.1.8 设函数 $f:(a,b) \to I\!R$, 那么

(1) 如果 f 在 (a,b) 上处处可微，则 f 是 (a,b) 上的凸函数当且仅当 $f'(x)$ 在 (a,b) 上是不减的；

(2) 如果 f 在 (a,b) 上处处二次可微，则 f 为 (a,b) 上的凸函数的充分必要条件是

$$f''(x) \geq 0, \quad \forall x \in (a,b).$$

一般说来，直接从定义来判断一个函数是否凸函数，往往不是一件容易的事情. 推论 2.1.8 用来判断一个光滑函数的凸性则是相当方便的. 此外，数学分析中的一些重要的不等式能够从一些函数的凸性推导出来，而这些不等式的直接

证明往往也是相当困难的. 例如，利用二阶导数，容易断定 $y=-\ln x$ 是 $(0,\infty)$ 上的凸函数. 于是由 Jensen 不等式可知， $\forall x_1>0,\cdots,x_n>0$,

$$\frac{1}{n}(\ln x_1+\cdots+\ln x_n)\leq\ln\frac{x_1+\cdots+x_n}{n},$$

即

$$(x_1\cdots x_n)^{1/n}\leq\frac{x_1+\cdots+x_n}{n},$$

这正是通常的几何平均值不超过算术平均值的公式.

函数的凸性可以与图象空间中的凸集联系起来. 对于函数 $f:(a,b)\to\mathbb{R}$, 我们称图象空间 $\mathbb{R}^2=\mathbb{R}\times\mathbb{R}$ 中的集合

$$\operatorname{epi}f=\{(x,y)\tau\mid x\in(a,b),f(x)\leq y\}$$

为 f 的上图 (epigraph).

定理 2.1.9 函数 $f:(a,b)\to\mathbb{R}$ 为凸函数的充分必要条件是: $\operatorname{epi}f$ 为 $\mathbb{R}^2$ 中的凸集.

证明 首先设 $f:(a,b)\to\mathbb{R}$ 是凸函数，而 $(x_1,y_1)^\tau$, $(x_2,y_2)^\tau\in\operatorname{epi}f$, 则 $x_1,x_2\in(a,b)$, $f(x_1)\leq y_1,f(x_2)\leq y_2$. 于是 $\forall\lambda\in[0,1]$, 从 f 的凸性我们得到

$$f(\lambda x_1+(1-\lambda)x_2)\leq\lambda f(x_1)+(1-\lambda)f(x_2)\leq\lambda y_1+(1-\lambda)y_2,$$

这表明

$$\begin{aligned}&\lambda(x_1,y_1)^\tau+(1-\lambda)(x_2,y_2)^\tau\\&=\Big(\lambda x_1+(1-\lambda)x_2,\lambda y_1+(1-\lambda)y_2\Big)^\tau\in\operatorname{epi}f.\end{aligned}$$

从而 $\operatorname{epi}f$ 是 $\mathbb{R}^2$ 中的凸集.

现在设 $\operatorname{epi}f$ 是 $\mathbb{R}^2$ 中的凸集， $x_1,x_2\in(a,b)$, $0\leq\lambda\leq1$, 注意到 $(x_1,f(x_1))^\tau,(x_2,f(x_2))^\tau\in\operatorname{epi}f$, 从 $\operatorname{epi}f$

的凸性我们得到

$$\left(\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda)f(x_2)\right)^{\tau}$$
$$= \lambda(x_1, f(x_1))^{\tau} + (1-\lambda)(x_2, f(x_2))^{\tau} \in \operatorname{epi} f.$$

由此从 $\operatorname{epi} f$ 的定义有

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda)f(x_2), \quad \forall \lambda \in [0,1].$$

这正是说 f 是 (a,b) 上的凸函数. ▮

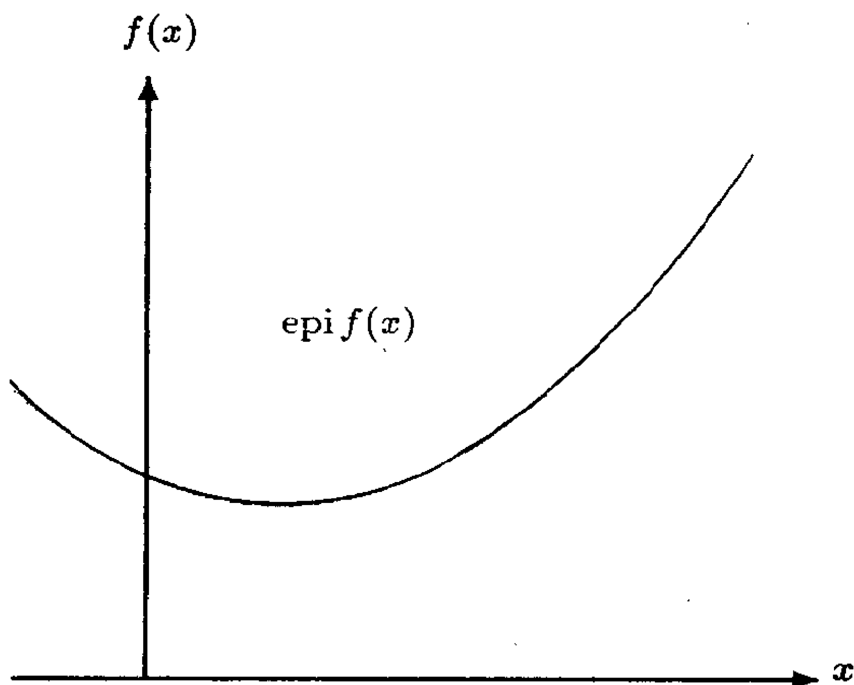


图 2.1.2　函数 f 的上图

设 $f:(a,b) \to \mathbb{R}$ 是凸函数, $x_0 \in (a,b)$. 如果 $\alpha \in \mathbb{R}$ 满足

$$f(x) \ge f(x_0) + \alpha(x - x_0), \quad \forall x \in (a,b), \tag{2.1.8}$$

则 α 叫做 f 在 x_0 处的次梯度. 次梯度的几何意义是: f

的上图 epi f 位于通过点 $(x_0, f(x_0))$ 的斜率为 α 的直线 $y = \alpha(x - x_0) + f(x_0)$ 的上方 (见图 2.1.3). 这可以看成导数概念的一种推广. f 在 x_0 处的次梯度的全体叫做 f 在 x_0 处的次微分，记作 $\partial f(x_0)$, 即

$$\partial f(x_0) = \{\alpha \mid f(x) \geq f(x_0) + \alpha(x - x_0),\ \forall\, x \in (a, b)\}.$$

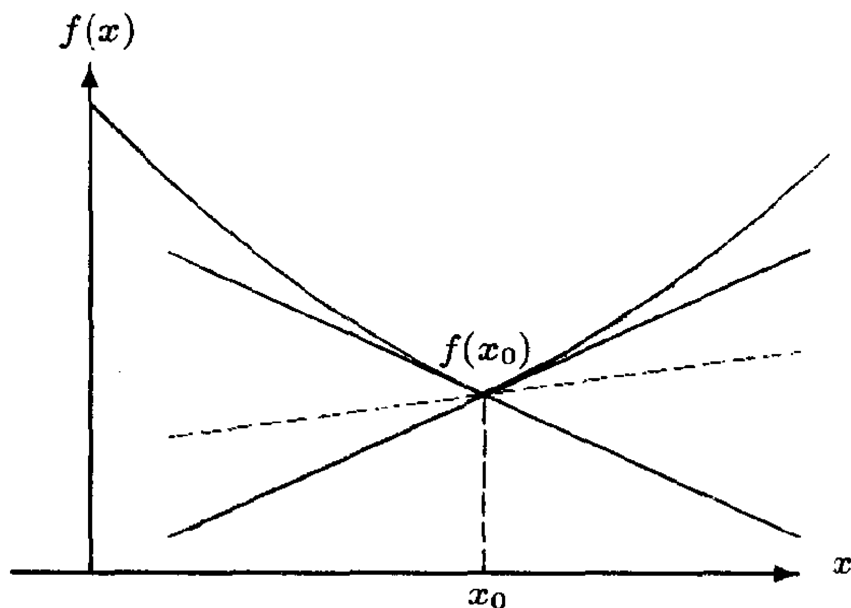


图 2.1.3　凸函数的次梯度和次微分

定理 2.1.10　设 $f : (a, b) \to I\!R$ 为凸函数，则

$$\partial f(x_0) = [f'_-(x_0); f'_+(x_0)], \quad \forall\, x_0 \in (a, b).$$

证明　设 $\alpha \in [f'_-(x_0), f'_+(x_0)]$, 则由 (2.1.4) 可得

$$\frac{f(x) - f(x_0)}{x - x_0} \geq f'_+(x_0) \geq \alpha, \quad \forall\, x > x_0;$$

$$\frac{f(x) - f(x_0)}{x - x_0} \leq f'_-(x_0) \leq \alpha, \quad \forall\, x < x_0.$$

综合上列两个不等式得到

$$\frac{f(x)-f(x_0)}{x-x_0} \geq \alpha, \quad \forall\, x > x_0;$$

$$\frac{f(x)-f(x_0)}{x-x_0} \leq \alpha, \quad \forall\, x < x_0.$$

由此即得 $f'_-(x_0) \leq \alpha \leq f'_+(x_0)$. ▮

例如，设

$$f(x) = |x|, \quad x \in \mathbb{R}.$$

那么

$$f'_+(x) = \begin{cases} +1, & x \geq 0, \\ -1, & x < 0, \end{cases} \quad f'_-(x) = \begin{cases} +1, & x > 0, \\ -1, & x \leq 0. \end{cases}$$

于是

$$\partial f(x) = \begin{cases} \{+1\}, & x > 0, \\ [-1,1], & x = 0, \\ \{-1\}, & x < 0. \end{cases}$$

习　题　2.1

2.1.1 试证：$2e^{x+y} \leq e^{2x} + e^{2y}$, $\forall\, x, y \in \mathbb{R}$.

2.1.2 试证：对于任意 $\alpha \geq 1$ 和任意非负实数 $x_1, \cdots, x_n$，下列不等式成立：

$$\left(\frac{x_1 + \cdots + x_n}{n}\right)^\alpha \leq \frac{x_1^\alpha + \cdots + x_n^\alpha}{n}.$$

2.1.3 试证下列函数是严格凸的：

(1) e^x 在 $(-\infty, \infty)$ 上；

(2) $-\ln x$ 在 $(0, \infty)$ 上；

(3) x^p 在 $[0, \infty)$ 上，这里 $p \geq 1$;

(4) $-x^p$ 在 $[0, \infty)$ 上，这里 $0 < p < 1$;

(5) $x \ln x$ 在 $[0, \infty)$ 上；

(6) $-\sin x$ 在 $[0,\pi]$ 上；

(7) $-\cos x$ 在 $[-\pi/2,\pi/2]$ 上；

(8) $\tan x$ 在 $[0,\pi/2)$ 上；

(9) $\arctan x$ 在 $(0,\pi/2]$ 上；

2.1.4 试证：在 Jensen 不等式 (2.1.2) 中若取 $\lambda_k = 1/m$, $k=1,\cdots,m$, 则等号当且仅当 $x_1=\cdots=x_m$ 时成立.

2.1.5 已知 a,b,c 是正实数，而且满足

$$\frac{a^2}{1+a^2}+\frac{b^2}{1+b^2}+\frac{c^2}{1+c^2}=1.$$

试证：$abc\le\sqrt{2}/4$.

2.1.6 证明 Hölder 不等式：设 a_i,b_i $(1\le i\le n)$ 为 $2n$ 个非负实数，$\alpha>0,\beta>0,\alpha+\beta=1$, 则

$$\sum_{i=1}^n a_i^\alpha b_i^\beta \le \left(\sum_{i=1}^n a_i\right)^\alpha \left(\sum_{i=1}^n b_i\right)^\beta.$$

2.1.7 设 $f:(a,b)\to\mathbb{R}$, 我们称 f 在 (a,b) 上是中间凸的，是指它满足

$$f\left(\frac{x+y}{2}\right)\le\frac{1}{2}\Big(f(x)+f(y)\Big),\quad \forall\, x,y\in(a,b).$$

求证：(1) f 在 (a,b) 上为中间凸的充分必要条件是，对于任意有理数 $\alpha_1\ge0,\cdots,\alpha_m\ge0$, $\alpha_1+\cdots+\alpha_m=1$, 有

$$f(\alpha_1x_1+\cdots+\alpha_mx_m)\le\alpha_1f(x_1)+\cdots+\alpha_mf(x_m),$$

$$\forall\, x,\cdots,x_m\in(a,b);$$

(2) 如果 f 在 (a,b) 上还是连续的，则 f 必是凸函数.

2.1.8 设 $[a,b]$ 为有穷区间，$[a,b]$ 上的凸函数 f 在区间 $[a,b]$ 的端点可以没有单侧导数，甚至可以不连续. 试证：f 在端点处必上半连续，即

$$\limsup_{x\to a+0}f(x)\le f(a),\quad \limsup_{x\to b-0}f(x)\le f(b).$$

2.1.9 设 f 是 $(0,\infty)$ 上的递增凸函数，求证：或者 f 为常值函数，或者

$$\lim_{x\to\infty} f(x) = +\infty.$$

2.1.10 设 f 是 $(0,\infty)$ 上的凸函数，试证：$\lim\limits_{x\to\infty} f(x)/x$ 存在 (可能等于 $+\infty$)，并且

$$\lim_{x\to\infty} f(x)/x = \lim_{x\to\infty} f'_+(x) = \lim_{x\to\infty} f'_-(x).$$

2.1.11 设 $f:(a,b)\to\infty$ 是连续函数，求证：为了 f 是 (a,b) 上的凸函数，必须且只须

$$\limsup_{h\to 0} \frac{f(x+h)+f(x-h)-2f(x)}{h^2} \geq 0, \quad \forall x\in(a,b).$$

2.1.12 设 f 是 (a,b) 上的凸函数，并且 $[a,b]\subset(0,\infty)$. 试证：为了 $f(1/x)$ 是 $(1/b,1/a)$ 上的凸函数，必须且只须 $xf(x)$ 是 (a,b) 上的凸函数.

2.1.13 设 $f:[0,2\pi]\to\mathbb{R}$ 为凸函数，$k\geq 1$ 为整数. 求证：

$$a_k = \int_0^{2\pi} f(x)\cos kx\, dx \geq 0.$$

2.1.14 设函数 $f:(a,b)\to\mathbb{R}$. 求证：为了 f 是 (a,b) 上的凸函数，必须且只须对于任意不同的三个数 $x_1,x_2,x_3\in(a,b)$，有

$$\frac{(x_3-x_2)f(x_1)+(x_1-x_3)f(x_2)+(x_2-x_1)f(x_3)}{(x_1-x_2)(x_2-x_3)(x_3-x_1)} \geq 0.$$

2.1.15 设 $f:(a,b)\to\mathbb{R}$ 为连续函数，试证 f 为凸函数的充分必要条件是：

$$\int_{x_1}^{x_2} f(x)dx\Big/(x_2-x_1) \leq \frac{1}{2}\Big[f(x_1)+f(x_2)\Big], \quad \forall a<x_1<x_2<b.$$

2.1.16 完成推论 2.1.8 的证明.

2.2 $\mathbb{R}^n$ 中凸函数的基本性质

现在我们讨论 $\mathbb{R}^n$ 中的凸函数. 设 K 是 $\mathbb{R}^n$ 中的函数，我们称 $f: K \to \mathbb{R}$ 为凸函数，是指它满足

$$\begin{cases} f((1-\lambda)x+\lambda y) \leq (1-\lambda)f(x)+\lambda f(y), \\ \qquad \forall\, x,y \in K,\ \forall\, 0 \leq \lambda \leq 1. \end{cases} \tag{2.2.1}$$

同单变量函数情形一样，我们有如下定理:

定理 2.2.1 (Jensen不等式) 为了凸集 $K \subset \mathbb{R}^n$ 上的函数 f 是凸函数，必须且只须如下 Jensen 不等式成立:

$$\begin{cases} f(\lambda_1 x_1 + \cdots + \lambda_m) \leq \lambda_1 f(x_1) + \cdots + \lambda_m f(x_m), \\ \qquad \forall\, x_i \in K, 0 \leq \lambda_i \leq 1,\ \lambda_1 + \cdots + \lambda_m = 1, \end{cases} \tag{2.2.2}$$

其中 m 是任意自然数.

证明 只需证明必要性，即证式 (2.2.1)$\Longrightarrow$ 式 (2.2.2). 仿照定理 2.1.1 的证明，我们采用数学归纳法. 当 $m=2$ 时式 (2.2.2) 正好是式 (2.2.1). 假定 $m=k$ 时式 (2.2.2) 成立 : $\forall\, x_1, \cdots, x_k \in K, \forall\, \lambda_i \geq 0, \lambda_1 + \cdots + \lambda_k = 1$, 我们有

$$f(\lambda_1 x_1 + \cdots + \lambda_k) \leq \lambda_1 f(x_1) + \cdots + \lambda_k f(x_k).$$

今对 $m=k+1$ 证明式 (2.2.2) 也成立. 为此任取 $x_1, \cdots, x_{k+1}$ 及 $\lambda_i \geq 0$, $\lambda_1 + \cdots + \lambda_{k+1} = 1$. 不妨假定 $0 < \lambda_{k+1} < 1$, 于是

$$0 < \mu \triangleq \lambda_1 + \cdots + \lambda_k < 1,\ \ 1-\mu = \lambda_{k+1}.$$

令

$$\mu_i \triangleq \lambda_i/\mu, \quad i=1,\cdots,k.$$

从 f 的凸性和归纳假设得到

$$\begin{aligned} f(\lambda_1 x_1 &+ \cdots + \lambda_{k+1}x_{k+1}) \\ &= f\Big(\mu(\mu_1 x_1 + \cdots + \mu_k x_k) + (1-\mu)x_{k+1}\Big) \\ &\le \mu f(\mu_1 x_1 + \cdots + \mu_k x_k) + (1-\mu)f(x_{k+1}) \\ &\le \mu\Big(\mu_1 f(x_1) + \cdots + \mu_k f(x_k)\Big) + \lambda_{k+1}f(x_{k+1}) \\ &= \lambda_1 f(x_1) + \cdots + \lambda_{k+1}x_{k+1}. \end{aligned}$$

▮

同样地我们可以把函数的凸性与图象空间中某个集合(上图) 的凸性联系起来.

对于函数 $f: K\subset \mathbb{R}^n \to \mathbb{R}$, 乘积空间 (即图象空间)$\mathbb{R}^{k+1}=\mathbb{R}^n\times\mathbb{R}$ 中的子集

$$\operatorname{epi} f \triangleq \{(x,\lambda)\in\mathbb{R}^{n+1}\mid x\in K,\ f(x)\le\lambda\}$$

称为 f 的上图. 仿照单变量情形，不难证明如下定理:

定理 2.2.2 设 $K\subset\mathbb{R}^n$ 为凸集，那么 $f: K\to\mathbb{R}$ 为凸函数的充分必要条件是: $\operatorname{epi} f$ 为 $\mathbb{R}^{n+1}$ 中的凸子集.

今后我们将允许函数 f 取广义实值 $\mathbb{R}\cup\{\pm\infty\}=[-\infty,\infty]$. 对于涉及 $\pm\infty$ 的运算，我们作如下的约定:

$$\alpha+\infty=\infty+\alpha=\infty,\ \forall -\infty<\alpha\le\infty;$$

$$\alpha-\infty=-\infty+\alpha=-\infty,\ \forall -\infty\le\alpha<\infty;$$

$$\alpha\infty=\infty\alpha=\infty,\ \alpha(-\infty)=(-\infty)\alpha=-\infty,\ \forall\,\alpha>0;$$

$$\alpha\infty=\infty\alpha=-\infty,\ \alpha(-\infty)=(-\infty)\alpha=\infty,\ \forall\,\alpha<0;$$

$$0\infty=\infty0=0(-\infty)=(-\infty)0=0,\ -(-\infty)=\infty;$$

$$\inf\emptyset=+\infty,\ \sup\emptyset=-\infty.$$

但 $\infty-\infty$ 或 $-\infty+\infty$ 则没有定义，应避免出现这种情形．对于 $f: K \to [-\infty,\infty]$，我们仍将把式 (2.2.1) 作为 f 凸性的定义，并且不难验证定理 2.2.1 和 2.2.2 对于 $f: K \to [-\infty,\infty]$ 仍然成立．

把凸函数的值域扩张到广义实值以后，任何凸集 $K \subset \mathbb{R}^n$ 上的凸函数 f 都可以用下列方式扩张到全空间 $\mathbb{R}^n$，成为 $\mathbb{R}^n$ 上的凸函数：

$$\tilde{f}(x)=\begin{cases} f(x), & x \in K, \\ +\infty, & x \in \mathbb{R}^n \setminus K. \end{cases}$$

这样，今后我们只要考虑全空间 $\mathbb{R}^n$ 上的广义实值凸函数就够了．

对于凸函数 $f: \mathbb{R}^n \to [-\infty,\infty]$，$\mathbb{R}^n$ 中的集合

$$\operatorname{dom} f=\{x \in \mathbb{R}^n \mid f(x)<\infty\}$$

叫做 f 的有效定义域；而当有效定义域 $\operatorname{dom} f \neq \emptyset$ 并且不含 $-\infty$ 时，f 叫做真凸函数．当 $\operatorname{dom} f=\emptyset$ 时，$f(x) \equiv \infty, \forall x \in \mathbb{R}^n$．因此真正有实际意义的凸函数是真凸函数．但从理论上考虑非真凸函数将使凸函数的理论更完整．注意不恒等于 $-\infty$ 的非真凸函数是存在的，例如在 $\mathbb{R}$ 上定义

$$f(x)=\begin{cases} -\infty, & \text{若 } |x|<1, \\ 0, & \text{若 } |x|=1, \\ +\infty, & \text{若 } |x|>1. \end{cases}$$

注意，对于 $f: \mathbb{R}^n \to [-\infty,\infty]$，把式 (2.2.1) 作为凸函数的定义有可能出现 $\infty-\infty$ 这样的不确定情况．但是用 f 的上图 $\operatorname{epi} f$ 的凸性来定义 f 的凸性就避免了这种困难．

下面我们列举一些 $\mathbb{R}^n$ 上的凸函数的例子．

例 2.2.1 $\mathbb{R}^n$ 上的仿射函数 f:

$$f(x) = \langle x, b\rangle + \beta, \quad \forall\, x \in \mathbb{R}^n,$$

这里 $b \in \mathbb{R}^n$ 和 $\beta \in \mathbb{R}$ 为给定的元和数. 显然仿射函数是凸的.

例 2.2.2 $\mathbb{R}^n$ 上的欧氏范数

$$\|x\| = (\xi_1^2 + \cdots + \xi_n)^{1/2}, \quad \forall\, x = (\xi_i) \in \mathbb{R}^n$$

是凸函数; 更一般地, $\mathbb{R}^n$ 上的 p 范数

$$\|x\|_p = (\xi_1^p + \cdots + \xi_n^p)^{1/p}, \quad \forall\, x = (\xi_i) \in \mathbb{R}^n$$

也是凸函数, 其中 $p \geq 1$.

例 2.2.3 给定集合 $K \subset \mathbb{R}^n$, K 的示性函数是指

$$\delta(x|K) = \begin{cases} 0, & x \in K, \\ +\infty, & x \in \mathbb{R}^n \backslash K. \end{cases}$$

不难看出, 当且仅当 K 为凸集时 $\delta(x|K)$ 是 (取广义实值的) 凸函数.

例 2.2.4 $\mathbb{R}^n$ 中凸集 K 的承托函数 $\sigma(x|K)$ 定义为

$$\sigma(x|K) = \sup\{\langle x, y\rangle \mid y \in K\};$$

K 的度规函数 (gauge) $\gamma(x|K)$ 定义为

$$\gamma(x|K) = \inf\{\lambda \geq 0 | x \in \lambda K\};$$

K 的 (欧几里得) 距离函数 $d(x|K)$ 定义为

$$d(x|K) = \inf\{\|x - y\| \mid y \in K\}.$$

这几个函数的凸性容易直接验证. 今后我们还将回过来讨论这些重要的函数. 下一个例子中讨论的重要的 Minkowski 函数是度规函数的特殊情形.

我们称函数 $f: \mathbb{R}^n \to (-\infty, \infty]$ 为次线性的，是指它满足

(1) $f(\lambda x) = \lambda f(x), \quad \forall\, x \in \mathbb{R}^n, \forall\, \lambda \geq 0$;(正齐性)

(2) $f(x+y) \leq f(x)+f(y), \quad \forall\, x, y \in \mathbb{R}^n$.(次加法性)

显然，次线性函数一定是凸函数，并且任何线性函数当然是次线性的；反之，如果 $f: \mathbb{R}^n \to \mathbb{R}$ 使得 f 和 $-f$ 都是次线性的，则 f 必是线性的.

例 2.2.5 设 $K \subset \mathbb{R}^n$ 是一有界闭凸集， $0 \in \operatorname{int} K$, 那么 K 的 Minkowski 函数

$$p_K(x) = \inf\{\alpha > 0 \mid x \in \alpha K\}, \quad \forall\, x \in \mathbb{R}^n$$

是 $\mathbb{R}^n$ 上处处取有穷值的次加法正齐性函数，即

$$p_K(x+y) \leq p_K(x) + p_K(y), \quad \forall\, x, y \in \mathbb{R}^n, \tag{2.2.3}$$

$$p_K(\lambda x) = \lambda p_K(x), \quad \forall\, x \in \mathbb{R}^n, \forall\, \lambda \geq 0, \tag{2.2.4}$$

从而 p_K 是 $\mathbb{R}^n$ 上的一个有穷值凸函数. 事实上，由于 $0 \in \operatorname{int} K$, f 在 $\mathbb{R}^n$ 上处处取有穷值是显然的. 设 $x \in \mathbb{R}^n$, $\lambda > 0$, 则

$$\begin{aligned} p_K(\lambda x) &= \inf\{\alpha > 0 \mid \lambda x \in \alpha K\} \\ &= \inf \lambda\{\alpha/\lambda > 0 \mid x \in (\alpha/\lambda)K\} \\ &= \lambda \inf\{\beta > 0 \mid x \in \beta K\} \\ &= \lambda p_K(x). \end{aligned}$$

现在任取 $x, y \in \mathbb{R}^n$, 依照 $p_K(x)$ 和 $p_K(y)$ 的定义，对于任意 $\varepsilon > 0$. 存在 $\lambda > 0$ 和 $\mu > 0$ 使得 $x \in \lambda K, y \in \mu K$, 并且

$$0 < \lambda < p_K(x) + \varepsilon/2, \quad 0 < \mu < p_K(y) + \varepsilon/2.$$

由 K 的凸性可得

$$\frac{x+y}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}\frac{x}{\lambda} + \frac{\mu}{\lambda+\mu}\frac{y}{\mu} \in K.$$

于是

$$p_K(x+y) \le \lambda + \mu < p_K(x) + p_K(y) + \varepsilon.$$

但 $\varepsilon >$ 是任意的，因此式 (2.2.3) 成立.

定理 2.2.3 设 I 为任一指标集， $\{f_i \mid i \in I\}$ 为 $I\!R^n$ 上的一族凸函数，那么

$$f(x) = \sup\{f_i(x) \mid i \in I\}, \quad \forall x \in I\!R^n \tag{2.2.5}$$

是 $I\!R^n$ 上的凸函数，记作 $f = \sup\{f_i \mid i \in I\}$.

证明 容易直接从凸函数的定义出发证明 f 的凸性，但最简单的证明方法是应用定理 2.2.2, 只要注意

$$\operatorname{epi} f = \cap\{\operatorname{epi} f_i \mid i \in I\},$$

这是因为，对于任意 $\mu \in I\!R$,

$$f(x) \le \mu \Longleftrightarrow f_i(x) \le \mu,\ \forall i \in I. \qquad \blacksquare$$

通常由式 (2.2.5) 定义的函数 f 叫做函数族 $\{f_i \mid i \in I\}$ 的上包络. 由此可见，一族仿射函数的上包络是凸函数. 实际上，我们可以证明，在一定条件下，凸函数可表示成某一仿射函数族的上包络. 我们将在下一节讨论这个问题.

给定函数 $f: I\!R^n \to [-\infty, \infty]$. 对于任意 $\alpha \in [-\infty, \infty]$, 集合

$$\{x \in I\!R^n \mid f(x) < \alpha\} \text{ 和 } \{x \in I\!R^n \mid f(x) \le \alpha\}$$

称为水平集.

定理 2.2.4 设 $f: I\!R^n \to [-\infty, \infty]$ 是凸函数. 那么对于一切 $\alpha \in [-\infty, \infty]$, 水平集 $\{x \in I\!R^n \mid f(x) < \alpha\}$ 和 $\{x \in I\!R^n \mid f(x) \le \alpha\}$ 是 $I\!R^n$ 中的凸集.

证明 留作练习. ∎

仿照单变量函数情形，现在我们来讨论 $\mathbb{R}^n$ 中的可微函数的凸性条件.

定理 2.2.5 设 $K \subset \mathbb{R}^n$ 为开凸集， $f : K \to \mathbb{R}$ 为可微函数，其导数记作 $f' = \left(\dfrac{\partial f}{\partial \xi_1}, \cdots, \dfrac{\partial f}{\partial \xi_n}\right)^\tau$. 那么 f 为 K 上凸函数的充分必要条件是:

$$\langle f'(x) - f'(y), x - y\rangle \geq 0, \quad \forall\, x, y \in\in K. \tag{2.2.6}$$

证明 由习题 2.2.3, 为了 $f : K \to \mathbb{R}$ 是凸函数，必须且只须对于任意固定的 $x, y \in K$,

$$\varphi_{x,y}(t) = f((1-t)x + ty)$$

作为 t 的函数在 $[0,1]$ 上是凸的；而且由定理 2.1.5, 为了 $\varphi_{x,y}(\cdot)$ 是凸函数，又必须且只须其导数 $\varphi'_{x,y}(t)$ 在 $[0,1]$ 上不减. 但在目前情况下,

$$\varphi'_{x,y}(t) = \langle f'(x + t(y-x)), y - x\rangle, \quad t \in [0,1].$$

如果 $\varphi'_{x,y}(t)$ 在 $[0,1]$ 上不减，则特别地有

$$\varphi'_{x,y}(t) \geq \varphi'_{x,y}(0),$$

这正好是式 (2.2.6). 反之，若式 (2.2.6) 成立，则对于任意 $t_1, t_2 \in [0,1], t_1 < t_2$, 从式 (2.2.6) 得到

$$\begin{aligned}
&\varphi'_{x,y}(t_2) - \varphi'_{x,y}(t_1) \\
&= \Big\langle f'(x + t_2(y-x)) - f'(x + t_1(y-x)), y - x\Big\rangle \\
&= \frac{1}{t_2 - t_1}\Big\langle f'(x + t_2(y-x)) - f'(x + t_1(y-x)), \\
&\quad x + t_2(y-x) - (x + t_1(y-x))\Big\rangle \geq 0,
\end{aligned}$$

这表明对于任意 $x,y\in K$, $\varphi'_{x,y}(t)$ 在 $[0,1]$ 上是不减的. ▮

我们称满足条件式 (2.2.6) 的映射 $f':\mathbb{R}^n\to\mathbb{R}^n$ 为单调映射. 我们将在第四章讨论凸函数的次微分时再回到这一问题.

定理 2.2.6 设 $K\subset\mathbb{R}^n$ 为开凸集, $f:K\to\mathbb{R}$ 为二次连续可微函数. 那么 f 为凸函数的充分必要条件是: 其海赛 (Hessian) 矩阵

$$A(x)=\Big(a_{ij}(x)\Big),\ a_{i,j}(x)=\frac{\partial^2 f}{\partial\xi_i\partial\xi_j}(x),\ x=(\xi_i)\in\mathbb{R}^n$$

对一切 $x\in K$ 为非负定矩阵, 即

$$\langle A(x)z,z\rangle\geq 0,\quad \forall\, x\in K,\ \forall\, z\in\mathbb{R}^n. \tag{2.2.7}$$

证明 仿照定理 2.2.5 证明中的推理, f 的凸性等价于对于任意固定的 $x,y\in\in K$, 函数 $\varphi_{x,y}(t)=f(x+t(y-x))$ 在 $[0,1]$ 上的凸性. 但在定理的假设下, $\varphi_{x,y}(t)$ 是 $[0,1]$ 上的二次连续可微函数, 因此, 依照定理 2.1.5, f 的凸性又等价于 $\varphi''_{x,y}(t)$ 在 $[0,1]$ 上的非负性, 即

$$\varphi''_{x,y}(t)\geq 0,\quad \forall\, t\in[0,1],\ \forall\, x,y\in K. \tag{2.2.8}$$

然而

$$\varphi''_{x,y}(t)=\Big\langle A(x+t(y-x))(y-x),y-x\Big\rangle.$$

由此由于式 (2.2.8) 中 $x,y\in K$ 和 $t\in[0,1]$ 的任意性以及 K 的开凸性, 即得式 (2.2.8) 等价于式 (2.2.7). ▮

例 2.2.6 设 A 是 $n\times n$ 对称实矩阵, $a\in\mathbb{R}^n$ 和 $\alpha\in\mathbb{R}$ 为给定向量和数. 从定理 2.2.6 可知: 二次函数

$$f(x)=\frac{1}{2}\langle x,Ax\rangle+\langle x,a\rangle+\alpha,\ x\in\mathbb{R}^n$$

为 $\mathbb{R}^n$ 上的凸函数的充分必要条件是 A 为非负定矩阵，这是因为 f 的 Hessian 矩阵恰好是 A.

例 2.2.7 考虑函数 $f:\mathbb{R}^n\to(-\infty,\infty]$,

$$f(x)=\begin{cases}-(\xi_1\cdots\xi_n)^{1/n}, & \forall\xi_i\geq 0,\\ +\infty, & \text{在别处}.\end{cases}$$

这里 $x=(\xi_1,\cdots,\xi_n)^\tau$. 直接计算得到，对于 $x=(\xi_1,\cdots,\xi_n)^\tau$, $y=(\eta_1,\cdots,\eta_n)^\tau\in\mathbb{R}^n$, 当 $\xi_1>0,\cdots,\xi_n>0$ 时,

$$\langle y,A(x)y\rangle=n^{-2}f(x)\left[\left(\sum_{j=1}^n\eta_j/\xi_j\right)^2-n\sum_{j=1}^n(\eta_j/\xi_j)^2\right],$$

这里 $A(x)$ 表示 f 的 Hessian 矩阵. 但 $f(x)<0$, 并且

$$(\alpha_1+\cdots+\alpha_n)^2\leq n(\alpha_1^2+\cdots+\alpha_n^2).$$

因此 $A(x)$ 是非负定的. 依照定理 2.2.6, f 是 $\mathbb{R}^n$ 上的凸函数.

定理 2.2.7 设 $f:\mathbb{R}^n\to(-\infty,\infty]$ 为正齐性函数. 那么 f 是凸函数的充分必要条件为

$$f(x+y)\leq f(x)+f(y),\quad\forall\,x,y\in\mathbb{R}^n.\tag{2.2.9}$$

证明 首先设 f 是凸函数. 于是对于任意 $x,y\in\mathbb{R}^n$, 利用 f 的正齐性和凸性得到

$$\begin{aligned}f(x+y)&=f(2(x/2+y/2))=2f(x/2+y/2)\\&\leq 2(f(x)/2+f(y)/2)=f(x)+f(y).\end{aligned}$$

反之，设式 (2.2.9) 成立，则对于任意 $x,y\in\mathbb{R}^n$, $0<\lambda<1$, 我们有

$$
\begin{aligned}
f((1-\lambda)x+\lambda y) &\le f((1-\lambda)x)+f(\lambda y) \\
&= (1-\lambda)f(x)+\lambda f(y),
\end{aligned}
$$

即 f 是凸函数. ▮

推论 2.2.8 设 f 是 $\mathbb{R}^n$ 上的正齐性真凸函数，则

$$
\begin{cases} f(\lambda_1 x_1+\cdots+\lambda_n x_n) \le \lambda_1 f(x_1)+\cdots+\lambda_n f(x_n), \\ \qquad \forall\, x_i \in \mathbb{R}^n,\ \lambda_i>0,\ 1\le i\le m. \end{cases}
$$

推论 2.2.9 设 f 是 $\mathbb{R}^n$ 上的正齐性真凸函数，则 $f(-x)\ge -f(x)$, $\forall\, x\in\mathbb{R}^n$.

证明 对于任意 $x\in\mathbb{R}^n$,

$$
f(x)+f(-x)\ge f(x-x)=f(0)\ge 0.
$$

▮

定理 2.2.10 设 f 是 $\mathbb{R}^n$ 上的正齐性真凸函数，L 是 $\mathbb{R}^n$ 的一个子空间. 那么为了 f 是 L 上的线性泛函，必须且只须

$$
f(-x)=-f(x), \quad \forall\, x\in L. \tag{2.2.10}
$$

又若 $\{b_1,\cdots,b_m\}$ 是 L 的基，则式 (2.2.10) 也等价于

$$
f(-b_i)=-f(b_i), \quad i=1,\cdots,m. \tag{2.2.11}
$$

证明 只需证明式 (2.2.11) 蕴涵 f 在 L 上的线性. 事实上，假定式 (2.2.11) 成立，则 $f(\lambda b_i)=\lambda f(b_i)$, $\forall\,\lambda\in\mathbb{R}$. 于是对于任意 $x=\lambda_1 b_1+\cdots+\lambda_m b_m\in L$, 我们有

$$
\begin{aligned}
f(\lambda_1 b_1)+\cdots+f(\lambda_m b_m) &\ge f(x)\ge -f(-x) \\
&\ge -\Big[f(-\lambda_1 b_1)+\cdots+f(-\lambda_m b_m)\Big] \\
&= f(\lambda_1 b_1)+\cdots+f(\lambda_m b_m).
\end{aligned}
$$

因此

$$\begin{aligned} f(x) &= f(\lambda_1 b_1) + \cdots + f(\lambda_m b_m) \\ &= \lambda_1 f(b_1) + \cdots + \lambda_m f(b_m), \end{aligned}$$

即 f 在 L 上是线性的. ∎

最后关于非真凸函数再作一些说明. 注意对于 $\mathbb{R}^n$ 上的非真凸函数 f, 其凸性条件式 (2.2.1) 应该用下列式子代替: 对于任意 $f(x) < \alpha$, $f(y) < \beta$, $0 < \lambda < 1$, 有

$$f((1-\lambda)x + \lambda y) < (1-\lambda)\alpha + \lambda\beta.$$

定理 2.2.11 设 f 是 $\mathbb{R}^n$ 上的一个非真凸函数, 那么

$$f(x) \equiv -\infty, \quad \forall\, x \in \mathrm{ri}\,(\mathrm{dom}\, f).$$

这就是说, 一个非真凸函数除了在其有效定义域的某些相对边界点外必定处处取无穷值.

证明 由于 f 是非真的, 则或者 $\mathrm{dom}\, f = \emptyset$, 或者至少有一点 $z \in \mathrm{dom}\, f$ 使得 $f(z) = -\infty$. 任取一点 $x \in \mathrm{ri}\,(\mathrm{dom}\, f)$. 根据定理 1.4.11, 存在 $\mu > 1$ 使得 $y = (1-\mu)z + \mu x \in \mathrm{dom}\, f$. 记 $\lambda = \mu^{-1}$, 则 $0 < \lambda < 1$, 并且 $x = (1-\lambda)z + \lambda y$. 于是对于任意 $\alpha > f(z), \beta > f(y)$, 有

$$f(x) = f((1-\lambda)z + \lambda y) < (1-\lambda)\alpha + \lambda\beta.$$

但 $f(z) = -\infty$, $f(y) < \infty$, 所以上式右端可以任意小, 因此 $f(x) = -\infty$. ∎

习 题 2.2

2.2.1 设 K 是 $\mathbb{R}^n$ 中的凸锥, $p \geq 1$ 为实数. 函数 $f: K \to \mathbb{R}$ 称为 p 次齐性的, 是指

$$f(\lambda x) = \lambda^p f(x), \quad \forall x \in K, \forall \lambda \geq 0.$$

假定 $f: K \to \mathbb{R}$ 是 p 次齐性凸函数，并且 $f(x) > 0, \forall x \in K\backslash\{0\}$，求证 $g(x) = [f(x)]^{1/p}$ 是 K 上的凸函数.

2.2.2 设 $\alpha_1, \cdots, \alpha_n$ 和 $\beta_1, \cdots, \beta_n$ 为非负实数，并设 $p \geq 1$. 求证 Minkowski 不等式

$$\left[\sum_{k=1}^{n}(\alpha_k + \beta_k)^p\right]^{1/p} \leq \left[\sum_{k=1}^{n}\alpha_k^p\right]^{1/p} + \left[\sum_{k=1}^{n}\beta_k^p\right]^{1/p}.$$

2.2.3 求证：为了 $f: K \to \mathbb{R}$ 是凸集 $K \subset \mathbb{R}^n$ 上的凸函数，必须且只须对于 K 上任意固定的两点 x, y，函数 $\varphi(t; x, y) \triangleq f(tx + (1-t)y)$ 是 $t \in [0,1]$ 的凸函数.

2.2.4 称函数 $f: \mathbb{R}^n \to \mathbb{R}$ 是拟凸的，是指对于一切 $\alpha \in \mathbb{R}$，水平集 $\{x \in \mathbb{R}^n \mid f(x) < \alpha\}$ 是 $\mathbb{R}^n$ 中的凸集. 试证：f 为拟凸的充分必要条件是

$$\begin{cases} f((1-\lambda)x + \lambda y) \leq \max\{f(x), f(y)\}, \\ \quad \forall x, y \in \mathbb{R}^n,\ \forall \lambda \in [0,1]. \end{cases}$$

2.2.5 设 $f: \mathbb{R}^n \to (-\infty, \infty]$ 为凸函数. 如果存在 $x_0 \in \operatorname{ri}(\operatorname{dom} f)$ 使得

$$f(x_0) = \sup\{f(x) \mid x \in \mathbb{R}^n\},$$

试证：f 为常值函数.

2.2.6 设 $K \subset \mathbb{R}^n$ 为凸集，$f: K \to \mathbb{R}$ 为凸函数，并且对于任意 $x \in K, f(x) \geq 0$. 求证 $f^2(x)$ 也是 K 上的凸函数，并用例子说明条件 $f(x) \geq 0$ 不能省略.

2.2.7 设 f 为 $\mathbb{R}^n$ 上正齐性函数. 试证：f 为凸函数当且仅当

$$f(x+y) \leq f(x) + f(y), \quad \forall x, y \in \mathbb{R}^n.$$

2.2.8 设 $g: \mathbb{R}^n \times \mathbb{R}^m \to (-\infty, \infty]$ 为真凸函数，令

$$f(x) = \inf\{g(x,y) \mid y \in \mathbb{R}^m\}.$$

试证：f 是 $\mathbb{R}^n$ 上的凸函数.

2.3 凸函数的运算

如何从已知的凸函数构造出新的凸函数当然也是一个十分有意思的问题. 同时我们将发现, 基于构造新的凸函数的这些原则, 反过来又可以为判别一些比较复杂的凸函数提供有用的方法.

一个凸函数左乘非负实数仍是凸函数, 两个凸函数之和也还是凸函数, 这些当然是凸函数运算的最简单的例子.

定理 2.3.1 设 f, f_1, f_2 是 $\mathbb{R}^n$ 上的真凸函数, 那么 $f_1 + f_2$ 是凸函数但未必是真凸的 (举例说明); 对于 $\alpha \geq 0$, αf 仍然是真凸的.

证明 留作练习. ∎

注意

$$(f_1 + f_2)(x) < \infty \text{ 等价于 } f_1(x) < \infty \text{ 和 } f_2(x) < \infty,$$

从而

$$\operatorname{dom}(f_1 + f_2) = (\operatorname{dom} f_1) \cap (\operatorname{dom} f_2).$$

定理 2.3.1 中之所以假定 f_1 和 f_2 是真凸函数, 目的是为避免出现 $\infty - \infty$ 这样的不确定情况.

定理 2.3.2 设 $f : \mathbb{R}^n \to (-\infty, \infty]$ 为凸函数, 并设 $\varphi : \mathbb{R} \to (-\infty, \infty]$ 为递增凸函数. 那么复合函数 $h(x) = \varphi(f(x))$ 也是 $\mathbb{R}^n$ 上的凸函数 (这里我们约定 $\varphi(+\infty) = +\infty$).

证明 对于 $x, y \in \mathbb{R}^n, \lambda \in [0, 1]$, 我们有

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y).$$

于是应用 φ 的单调递增性和凸性，得到

$$\begin{aligned} h((1-\lambda)x+\lambda y) &= \varphi\Big(f((1-\lambda)x+\lambda y)\Big) \\ &\leq \varphi\Big((1-\lambda)f(x)+\lambda f(y)\Big) \leq (1-\lambda)h(x)+\lambda h(x), \end{aligned}$$

即 h 是凸函数. ∎

从定理 2.3.2 知，如果 f 是 $\mathbb{R}^n$ 上的真凸函数，则 $e^{f(x)}$ 也是 $\mathbb{R}^n$ 上的真凸函数；如果 $f:\mathbb{R}^n\to\mathbb{R}_+$ 是非负凸函数，则 $g(x)\triangleq f(x)^p$ 也是 $\mathbb{R}^n$ 上的凸函数，这里 $p\geq 1$. 事实上，只需注意

$$\varphi(\xi)=\begin{cases}\xi^p, & \text{当 } \xi\geq 0,\\ 0, & \text{当 } \xi<0,\end{cases}$$

是 $\mathbb{R}\to\mathbb{R}$ 的递增函数. 特别由此可知当 $p\geq 1$ 时 $h(x)\triangleq\|x\|^p$ 是 $\mathbb{R}^n$ 上的凸函数，这里 $\|\cdot\|$ 是 $\mathbb{R}^n$ 上的欧氏范数.

如果 f 是 $\mathbb{R}^n$ 上的有限凸函数 (指 $f(x)$ 在 $\mathbb{R}^n$ 上处处取有穷值的凸函数), 而 K 是 $\mathbb{R}^n$ 的一个非空凸集，则

$$f(x)+\delta(x|K)=\begin{cases}f(x), & \text{当 } x\in K,\\ +\infty, & \text{当 } x\in\mathbb{R}^n\backslash K,\end{cases}$$

这里 $\delta(\cdot|K)$ 为 K 的示性函数. 因此，f 加上示性函数相当于限制 f 的有效定义域.

定理 2.3.3 设 F 是 $\mathbb{R}^{n+1}$ 中的一个凸子集，并设

$$f(x)=\inf\{\mu\mid(x,\mu)\in F\},\quad \forall x\in\mathbb{R}^n,$$

则 f 是 $\mathbb{R}^n$ 上的凸函数. 这里以及今后，我们约定对于实数域中的空集 $\emptyset$,

$$\inf\emptyset=+\infty,\quad \sup\emptyset=-\infty.$$

证明 只需注意，由于 F 的凸性，从 $(x,\alpha)\in F,(y,\beta)\in F$ 可得

$$(\lambda x+(1-\lambda y,\lambda\alpha+(1-\lambda)\beta)\in F,\quad \forall 0\le\lambda\le 1.$$ ∎

直观上，定理 2.3.3 中由凸集 $F\subset \mathbb{R}^{n+1}$ 决定的函数 f 是 F 的"下边界".

定理 2.3.4 设 $f_1,\cdots,f_m$ 是 $\mathbb{R}^n$ 上的真凸函数，对于 $x\in\mathbb{R}^n$，令

$$f(x)=\inf\left\{\sum_{k=1}^{m}f_k(x_k)\middle|x_k\in\mathbb{R}^n,\sum_{k=1}^{m}x_k=x\right\},$$

则 f 是 $\mathbb{R}^n$ 上的凸函数. 这里下确界是对一切满足 $x=x_1+\cdots+x_m$ 的 x_k 取的.

证明 设 $F_k=\operatorname{epi}f_k$，而 $F=F_1+\cdots+F_m$. 显然 F 是 $\mathbb{R}^{n+1}$ 中的凸集. 根据定义，$(x,\mu)\in F$ 当且仅当存在 $x_k\in\mathbb{R}^n,\mu_k\in\mathbb{R}$ 使得 $f_k(x_k)\le\mu_k$，并且 $(x,\mu)=(x_1,\mu_1)+\cdots+(x_m,\mu_m)$，即 $x=x_1+\cdots+x_m$，$\mu=\mu_1+\cdots+\mu_m$. 由此可见，

$$f(x)=\inf\{\mu\mid(x,\mu)\in F\},$$

从而依据定理 2.3.3，f 是凸函数. ∎

定理 2.3.4 中的凸函数 f 记作 $f_1\Box f_2\cdots\Box f_m$，叫做 $f_1,\cdots,f_m$ 的下端卷积. 特别当仅涉及两个函数 f 和 g 时，

$$(f\Box g)(x)=\inf\{f(x-y)+g(y)\mid y\in\mathbb{R}^n\},$$

这有些类似于经典的积分卷积公式.

如果 $g=\delta(\cdot|a)$ 是单点集 $\{a\}$ 的示性函数，则

$$\operatorname{dom}(f\Box g)(x)=f(x-a).$$

一般地，$f\Box g$ 的有效定义域是 f 和 g 的有效定义域之和，即

$$\operatorname{dom}(f\Box g)=\operatorname{dom}f+\operatorname{dom}g.$$

现在我们来定义凸函数的右乘标量运算. 设 f 是 $\mathbb{R}^n$ 上一凸函数，λ 是非负实数，$0\le\lambda<\infty$. 我们定义 $f\lambda$ 为凸函数

$$(f\lambda)(x)=\inf\{\mu\mid(x,\mu)\in\lambda\operatorname{epi}f\},\quad \lambda\ge 0,\, x\in\mathbb{R}^n.$$

于是

$$(f\lambda)(x)=\lambda f(\lambda^{-1}x),\quad \lambda>0;$$

当 $\lambda=0$ 并且 f 不恒等于 $+\infty$ 时，

$$(f0)(x)=\delta(x|0);$$

而当 $f\equiv+\infty$ 时，显然 $f0=f$. 函数 f 为正齐性的充分必要条件是 $f\lambda=f,\ \forall\lambda>0$.

设 h 是 $\mathbb{R}^n$ 中的任一凸函数；F 是由 $\operatorname{epi}h$ 产生的 $\mathbb{R}^{n+1}$ 中的凸锥. 于是按照定理 2.3.3, 从凸锥 F 所得到的函数 f 是正齐性凸函数. 注意 F 是 $\mathbb{R}^{n+1}$ 的原点与所有集合 $\lambda(\operatorname{epi}h)$ $(\lambda>0)$ 之并. 从而

$$f(x)=\inf\{(h\lambda)(x)\mid\lambda\ge 0\},\quad x\in\mathbb{R}^n.$$

今若 $g(x)$ 是 $\mathbb{R}^n$ 上任一正齐性凸函数，使得 $g\le h$ 并且 $g(0)\le 0$, 则 $g(x)\le(h\lambda)(x),\ \forall\lambda>0$. 由于 $(h0)(x)=\delta(x|0)$, 可知 $g\le f$, 故 f 是上述这样的正齐性凸函数中最大的一个.

设 f 是 $\mathbb{R}^n$ 上的真凸函数，则

$$g(\lambda,x)=\begin{cases}(f\lambda)(x), & 若\ \lambda\ge 0,\\ +\infty, & 若\ \lambda<0\end{cases}$$

是 $\mathbb{R}^{n+1}$ 上的正齐性真凸函数，这是由函数

$$h(\lambda,x)=\begin{cases}f(x), & \lambda=1,\\ +\infty, & \lambda\neq 1\end{cases}$$

所生成的正齐性凸函数. 特别 $\varphi(\lambda)=(f\lambda)(x)$ 对于任意固定的 $x\in\operatorname{dom}f$ 是 $\lambda\geq 0$ 的真凸函数.

$\mathbb{R}^n$ 中非空凸集 M 的度规函数 $\gamma(\cdot|M)$ 是由 $\delta(\cdot|M)+1$ 所生成的正齐性凸函数. 事实上，令 $h(x)=\delta(x|M)+1$, 我们有 $(h\lambda)(x)=\delta(x|\lambda M)+\lambda$. 于是,

$$\inf\{(h\lambda)(x)\mid\lambda\geq 0\}=\inf\{\lambda\geq 0\mid x\in\lambda M\}=\gamma(x|M).$$

一个非凸函数 g 的凸包 $\operatorname{conv}g$ 是指根据定理 2.3.3 从凸集 $F=\operatorname{conv}(\operatorname{epi}g)$ 所得到的函数 f, 它是不超过 g 的最大的凸函数. 根据定理 1.3.4, $(x,\mu)\in F$ 可以表示成凸组合

$$(x,\mu)=\sum_{k=1}^{m}\lambda_k(x_k,\mu_k)=\left(\sum_{k=1}^{m}\lambda_k x_k,\sum_{k=1}^{m}\lambda_k\mu_k\right),$$

其中 $(x_k,\mu_k)\in\operatorname{epi}g$, 即 $g(x_k)\leq\mu_k$. 于是对于 $x\in\mathbb{R}^n$,

$$(\operatorname{conv}g)(x)=\inf\left\{\sum_{k=1}^{m}\lambda_k g(x_k)\;\middle|\;\sum_{k=1}^{m}\lambda_k x_k=x\right\},\quad(2.3.1)$$

这里下确界是在 x 的所有这种凸组合形式上取的，当然要求 g 不取 $-\infty$, 以避免求和发生困难.

$\mathbb{R}^n$ 上任意一族函数 $\{f_i\mid i\in I\}$ 的凸包是指这一族函数的点点下确界函数的凸包，记作 $\operatorname{conv}\{f_i\mid i\in I\}$, 它是 $\mathbb{R}^n$ 上不超过每个 f_i 的最大的凸函数.

定理 2.3.5 设 $\{f_i\mid i\in I\}$ 是 $\mathbb{R}^n$ 上的一族真凸函数

(I 为任一指标集), 并设 $f = \operatorname{conv}\{f_i \mid i \in I\}$. 那么

$$f(x) = \inf\left\{ \sum_{i\in I} \lambda_i f(x_i) \,\middle|\, \sum_{i\in I} \lambda_i x_i = x \right\},\ x \in \mathbb{R}^n, \tag{2.3.2}$$

这里下确界是对 x 的一切元 x_i 的有穷凸组合形式取的 (即凸组合系数 λ_i 仅有穷个不为零). 式 (2.3.2) 当 x_i 限制在 $\operatorname{dom} f_i$ 中时也成立.

证明 如果记 $F = \operatorname{conv}(\cup\{\operatorname{epi} f_i \mid i \in I\})$, 则

$$f(x) = \inf\{\mu \mid (x,\mu) \in F\}.$$

根据定理 1.3.4, $(x,\mu) \in F$ 可以表示成如下有穷凸组合形式:

$$(x,\mu) = \sum_{i\in I} \lambda_i (x_i,\mu_i) = \left(\sum_{i\in I} \lambda_i x_i, \sum_{i\in I} \lambda_i \mu_i \right),$$

其中 $(x_i,\mu_i) \in \operatorname{epi} f_i$. 于是 $f(x)$ 是 $\sum_{i\in I} \lambda_i \mu_i$ 在 x 表示成一切有穷凸组合形式 $x = \sum_{i\in I} \lambda_i x_i$ (满足 $f_i(x_i) \le \mu_i$) 上取的下确界, 即式 (2.3.2) 成立. ∎

注 实际上, 公式 (2.3.2) 中的下确界只要取在所有这样一些凸组合上, 这些凸组合中至多 $n+1$ 个系数 λ_i 不为零, 并且相应的 x_i 是仿射无关的. 这可以应用 Caratheodory 定理来证明. 在目前情况下, $f(x)$ 是 $\mathbb{R}^{n+1}$ 中凸集 $F = \operatorname{conv}(\cup\{\operatorname{epi} f_i \mid i \in I\})$ 的下边界. 但从 Caratheodory 定理可知, F 中每个点可表示成不超过 $n+1$ 个仿射无关点 x_i 的凸组合, 这里每个 x_i 属于不同的 $\operatorname{epi} f_i$. 于是 $\mu > f(x)$ 当且仅当存在某个 $\alpha < \mu$ 使得 (x,α) 属于某个单纯形 S, 其顶点分属不同的 $\operatorname{epi} f_i$. 令

$$\alpha_0 = \min\{\alpha' \mid (x,\alpha') \in S\}.$$

显然 $(x,\alpha_0) \in S$. 这样 $(x,\alpha_0) \in S$ 可以表示成 S 的较少顶点的凸组合，这些顶点生成一个“子单纯形”，它不含任何“垂直”线段. 这个子单纯形的顶点记作 $(z_1,\alpha_1),\cdots,(z_m,\alpha_m)$. 容易验证：$y_1,\cdots,y_m$ 本身也是仿射无关的. 因此，$f(x)$ 是使得 (x,α) 可以表示成 $(y_1,\alpha_1),\cdots,(y_m,\alpha_m)$ 的凸组合的所有这样的 α 的下确界，这里 y_i 属于不同的 $\operatorname{epi} f_i$ 并且仿射无关. 但 $y_1,\cdots,y_m$ 仿射无关意味着 $m \le n+1$，由此即得所需要的结论.

特别是若

$$f_i(x) = \delta(x|a_i) + \alpha_i, \quad i \in I,$$

其中 $a_i \in \mathbb{R}^n, \alpha_i \mathbb{R}$ 为给定的向量和数，$\delta(\cdot|a_i)$ 为单点集 $\{a_i\}$ 的示性函数，则 $\operatorname{conv}\{f_i \mid i \in I\}$ 是满足

$$f(a_i) \le \alpha_i, \quad i \in I$$

的最大的凸函数，并且，

$$f(x) = \inf\left\{\sum_{i\in I}\lambda_i\alpha_i \,\middle|\, \sum_{i\in I}\lambda_i a_i = x\right\},$$

这里 $\sum\limits_{i\in I}\lambda_i\alpha_i$ 表示诸 a_i 的有穷凸组合.

定理 2.3.5 中的函数 f 也可以用下端卷积表示出来. 比如说，指标集 I 为有穷集 $I=\{1,\cdots,m\}$，而 $f_1,\cdots,f_m$ 是 $\mathbb{R}^n$ 上的真凸函数，则 f 可以通过定理 2.3.3 从凸集

$$\begin{aligned} F &= \operatorname{conv}\{\operatorname{epi} f_1,\cdots,\operatorname{epi} f_m\} \\ &= \cup\{\lambda_1 \operatorname{epi} f_1 + \cdots + \lambda_m \operatorname{epi} f_m\} \end{aligned}$$

得出，这里集合的并运算是对集合 epi f_k 是一切凸组合取的. 不难看出，下端卷积 $f_1\lambda_1\square\cdots\square f_m\lambda_m$ 是根据定理 2.3.3 从 $\mathbb{R}^{n+1}$ 中的凸集 $\lambda_1\text{epi}\, f_1+\cdots+\lambda_1\text{epi}\, f_m$ 所得到的函数. 由此可见， $f=\text{conv}\,\{f_1,\cdots,f_m\}$ 是

$$f(x)=\inf\left\{(f_1\lambda_1\square\cdots\square f_m\lambda_m)(x)\;\middle|\;\lambda_i\geq 0,\sum_{i=1}^{m}\lambda_i=1\right\}.$$

定理 2.3.6 设 $A:\mathbb{R}^n\to\mathbb{R}^m$ 是一个线性变换. 那么对于 $\mathbb{R}^m$ 上的每一个凸函数 g，由

$$(gA)(x)=g(Ax),\quad\forall\, x\in\mathbb{R}^n$$

所定义的函数 gA 是 $\mathbb{R}^n$ 上的凸函数；而对于 $\mathbb{R}^n$ 上的每一个凸函数 h，由

$$(Ah)(y)=\inf\{h(x)\mid Ax=y\}$$

所定义的函数 Ah 是 $\mathbb{R}^m$ 上的凸函数.

证明 直接从定义出发验证，留作练习. ▮

定理 2.3.6 中的函数 Ah 叫做 h 在 A 之下的象，而 gA 则叫做 g 在 A 之下的逆象. 容易看出，当 $A:\mathbb{R}^n\to\mathbb{R}^m$ 为非奇异变换时， $Ah=hA^{-1}$.

现在我们利用回收锥来给出凸函数的另一种运算. 设 f 是 $\mathbb{R}^n$ 上的不恒等于 $+\infty$ 的凸函数. f 的上图 epi f 是 $\mathbb{R}^{n+1}$ 中的一个非空凸集，因此可以考虑其回收锥 $0^+(\text{epi}\, f)$. 根据定义， $(y,\nu)\in 0^+(\text{epi}\, f)$ 当且仅当对于一切 $(x,\mu)\in$ epi f 和 $\lambda\geq 0$ 有

$$(x,\mu)+\lambda(y,\nu)=(x+\lambda y,\mu+\lambda\nu)\in\text{epi}\, f.$$

上式等价于

$$f(x+\lambda y) \le f(x)+\lambda\nu, \quad \forall\, x \in \operatorname{dom} f,\ \lambda \ge 0. \quad (2.3.3)$$

实际上，我们知道式 (2.3.3) 对于 $\lambda = 1$ 能推出它对一切 $\lambda \ge 0$ 成立. 对于给定的 y, 使得 $(y,\nu) \in 0^+(\operatorname{epi} f)$ 成立的 ν 值形成 $\mathbb{R}$ 中的一个上无界的闭区间. 因此， $0^+(\operatorname{epi} f)$ 是某个凸函数的上图. 我们称此凸函数为 f 的回收函数 (recession function), 记作 $f0^+$. 于是

$$\operatorname{epi}(f0^+) = 0^+(\operatorname{epi} f).$$

定理 2.3.7 设 f 为 $\mathbb{R}^n$ 上的真凸函数，那么 f 的回收函数 $f0^+$ 是正齐性真凸函数，并且对于每一 $y \in \mathbb{R}^n$, 有

$$(f0^+)(y) = \sup\{f(x+y) - f(x) \mid x \in \operatorname{dom} f\}. \quad (2.3.4)$$

如果 $\operatorname{epi} f$ 还是闭的，则 $\operatorname{epi}(f0^+)$ 也是闭的，并且对任意 $x \in \operatorname{dom} f$, 有

$$\begin{aligned}(f0^+)(y) &= \sup_{\lambda>0} \frac{f(x+\lambda y)-f(x)}{\lambda} \\ &= \lim_{\lambda\to\infty} \frac{f(x+\lambda y)-f(x)}{\lambda}. \quad (2.3.5)\end{aligned}$$

证明 式 (2.3.4) 是前面说明的直接结论. 从式 (2.3.3) 可知: $\nu \ge (f0^+)(y)$ 意味着

$$\nu \ge \sup_{\lambda>0} \frac{f(x+\lambda y)-f(x)}{\lambda}, \quad \forall\, x \in \operatorname{dom} f. \quad (2.3.6)$$

由此可见， $(f0^+)(y)$ 不可能取 $-\infty$. 对于任意固定的 $x \in \operatorname{epi} f$, 在 $\operatorname{epi} f$ 所含的从 $(x, f(x))$ 出发沿所有 (y,ν) 方向的半直线中，式 (2.3.6) 的上确界给出最小的实数 ν. 如果 $\operatorname{epi} f$ 是闭的，则依定理 1.8.1, 这个最小的 ν 值与 $x \in \operatorname{dom} f$

无关. 再考虑到差商 $(f(x+\lambda y)-f(x))/\lambda$ 是 λ 的递增函数，这样就证明了式 (2.3.5). 上图 $0^+(\mathrm{epi}\, f)$ 是非空凸锥，当 $\mathrm{epi}\, f$ 是闭的时候它也是闭的；因此 $f0^+$ 是一个正齐性的真凸函数，而且当 f 的上图 $\mathrm{epi}\, f$ 是闭的时候， $f0^+$ 的上图 $\mathrm{epi}\,(f0^+)$ 也是闭的. ∎

对于 $\mathbb{R}^n$ 上的凸函数 f, 当向量 $y\in\mathbb{R}^n$ 满足 $(f0^+)(y)\le 0$ 时， y 称为 f 的回收方向， f 的回收方向全体是一个凸锥，叫做 f 的回收锥 (注意不要与 $\mathrm{epi}\, f$ 的回收锥相混淆).

定理 2.3.8 设 f 为 $\mathbb{R}^n$ 上的真凸函数， $y\in\mathbb{R}^n$. 那么为了 y 是 f 的回收方向，即 $(f0^+)(y)\le 0$, 必须且只须对于一切 $x\in\mathbb{R}^n$, $f(x+\lambda y)$ 作为 $\lambda\in\mathbb{R}$ 的函数是 $\mathbb{R}$ 上的凸函数.

证明 根据回收函数 $f0^+$ 的定义， $(f0^+)(y)\le 0$ 当且仅当 $\mathrm{epi}\, f$ 的回收锥含有向量 $(y,0)$, 这意味着

$$f(z+\lambda y)\le f(z),\quad \forall\, z\in\mathbb{R}^n,\ \forall\,\lambda\ge 0. \tag{2.3.7}$$

但式 (2.3.7) 正是说对于任意 $x\in\mathbb{R}^n$, $f(x+\lambda y)$ 作为 $\lambda\in\mathbb{R}$ 的函数是非增的. ∎

在结束本节之前，我们对有关凸函数的几种主要的运算作一简单的小结：

(1) λf, $f+g$— 凸函数的加法和左乘标量.

(2) $f\lambda$— 凸函数右乘标量.

(3) $f0^+$— 凸函数的回收函数.

(4) $\mathrm{conv}\, g$— 非凸函数的凸包.

(5) $\sup\{f_i \mid i\in I\}$— 凸函数族的上包络.

(6) $f_1\Box\cdots\Box f_m$— 凸函数的下端卷积.

(7) $\operatorname{conv}\{f_i \mid i \in I\}$— 函数族的凸包.

(8) Af, gA— 凸函数左乘和右乘线性变换.

习　题　2.3

2.3.1 设 f, g 是 $\mathbb{R}^n$ 上的真凸函数，试证：$f+g$ 也是凸函数，但未必是真凸的，举例说明之；对于一切 $\lambda \geq 0$, λf 也是真凸的.

2.3.2 设 $f(x) = \max\{\xi_i \mid 1 \leq i \leq n\}, \forall x = (\xi_1, \cdots, \xi_n)^\tau \in \mathbb{R}^n$. 求证：$f$ 是真凸函数，并且如果 M 是 n 维单纯形

$$M = \{(\xi_1, \cdots, \xi_n)^\tau \mid \xi_i \geq 0, \xi_1 + \cdots + \xi_n = 1\},$$

则

$$f(x) = \sup\{\langle x, x^* \rangle \mid x^* \in M\}, \ \forall x \in \mathbb{R}^n.$$

2.3.3 设 $Q: \mathbb{R}^n \to \mathbb{R}^n$ 为正定对称矩阵，并设凸函数 $f(x) = \frac{1}{2}\langle Qx, x \rangle$. 试证：

$$(f0^+)(y) = \begin{cases} 0, & y = 0, \\ +\infty, & y \neq 0. \end{cases}$$

2.3.4 设 $f: \mathbb{R}^n \to (-\infty, \infty]$ 是凸函数. 如果 f 的上图 $\operatorname{epi} f$ 是 $\mathbb{R}^{n+1}$ 中的一个凸锥，那么 f 是正齐性的.

2.3.5 设 M 是 $\mathbb{R}^n$ 中的非空闭有界集. 设 f 是 M 上的连续实值函数，并设 $f(x) = +\infty, \forall x \notin M$. 试证：$\operatorname{conv} f$ 是 $\mathbb{R}^n$ 上的闭真凸函数.

2.3.6 设 f 为 $\mathbb{R}^n$ 上的仿射函数，$f(x) = \langle x, x^* \rangle + \alpha$. 试证：$(f0^+)(x) = \langle x, x^* \rangle$.

2.4 凸函数的下半连续性

$\mathbb{R}^n$ 上的线性函数的连续性是线性的简单结论. 在 2.1 节我们已看到，定义在开区间上的有穷值单变量凸函数也是处处连续的. 然而对于 $\mathbb{R}^n$ 上的广义实值凸函数，事情远非

如此简单. 但函数的凸性毕竟蕴涵着许多拓扑性质，其主要结论之一是：对于凸函数而言，下半连续性是一种“构造性”的性质，即存在一种简单的运算，只要对凸函数的有效定义域上某些边界点处重新赋值，就能使该凸函数具有下半连续性.

定理 2.4.1 设 $f: \mathbb{R} \to [-\infty, \infty]$, 则下列三个命题彼此等价：

(1) f 在 $\mathbb{R}^n$ 上是下半连续的；

(2) $\forall \alpha \in \mathbb{R}$, 水平集 $S_\alpha(f) = \{x \in \mathbb{R}^n \mid f(x) \le \alpha\}$ 是闭集；

(3) f 的上图 $\operatorname{epi} f$ 是 $\mathbb{R}^{n+1}$ 中的闭集.

证明 注意 f 在 x 处下半连续等价于

$$\begin{cases} \forall (x_k) \subset \mathbb{R}^n,\ \forall (\mu_k) \subset \mathbb{R},\ \mu_k \ge f(x_k),\ k \ge 1, \\ x = \lim\limits_{k\to\infty} x_k,\ \mu = \lim\limits_{k\to\infty} \mu_k \Longrightarrow \mu \ge f(x). \end{cases} \tag{2.4.1}$$

但式 (2.4.1) 正好意味着上图 $\operatorname{epi} f$ 的闭性，从而 (1) 与 (3) 等价. 另一方面，如果在式 (2.4.1) 中取 $\alpha = \mu = \mu_1 = \mu_2 = \cdots$, 则立即从 f 的下半连续性推出水平集 $S_\alpha(f)$ 的闭性，即 (1) $\Longrightarrow$ (2). 最后设 (2) 成立，并任意取 $(x_k) \subset \mathbb{R}^n$ 使得 $\lim\limits_{k\to\infty} x_k = x, \lim\limits_{k\to\infty} f(x_k) = \mu$. 显然对于任意 $\alpha > \mu$, 当 k 充分大后 $x_k \in S_\alpha(f)$, 于是从水平集 $S_\alpha(f)$ 的闭性知 $x \in S_\alpha(f), \forall \alpha > \mu$. 这表明 $f(x) \le \mu$, 而由上述 (x_k) 取法的任意性知

$$\liminf_{y\to x} f(y) \ge f(x),$$

即 f 下半连续. ∎

给定 $\mathbb{R}^n$ 的一个上凸函数 f, 存在一个不超过 f 的 (不一定取有穷值的) 最大的下半连续函数，并且不难看出这个函数的上图正好是 f 的上图的闭包. 通常我们称此函数为 f 的下半连续包.

凸函数 $f:\mathbb{R}^n\to(-\infty,\infty]$ 的闭包定义作 f 的下半连续包；而当 $f:\mathbb{R}^n\to[-\infty,\infty]$ 是非真的凸函数并且对某些 x 有 $f(x)=-\infty$ 时， f 的闭包定义作常值函数 $-\infty$. f 的闭包记作 $\mathrm{cl}\, f$.

$\mathbb{R}^n$ 上的凸函数 f 叫做闭的，是指它满足 $\mathrm{cl}\, f=f$. 这样，对于一个真凸函数，其闭性等价于其下半连续性.

给定 $\mathbb{R}^n$ 上的一个真凸函数 f, 从定义有

$$\mathrm{epi}\,(\mathrm{cl}\, f)=\mathrm{cl}\,(\mathrm{epi}\, f),$$

而且

$$(\mathrm{cl}\, f)(x)=\inf\{\mu\in\mathbb{R}\mid(x,\mu)\in\mathrm{cl}\,(\mathrm{epi}\, f)\},\ \ x\in\mathbb{R}^n.$$

由此可见

$$(\mathrm{cl}\, f)(x)=\liminf_{y\to x} f(y),\ \ x\in\mathbb{R}^n.$$

直观上， $\mathrm{cl}\, f$ 是由凸集 $\mathrm{cl}\,(\mathrm{epi}\, f)$ 的下边界所界定的凸函数. 下面我们列举几个简单的闭包运算的例子.

例 2.4.1 考虑凸函数 $f:\mathbb{R}\to(-\infty,\infty]$

$$f(x)=\begin{cases}0, & x>0,\\ \infty, & x\le 0.\end{cases}$$

那么

$$(\mathrm{cl}\, f)(x)=\begin{cases}0, & x\ge 0,\\ \infty, & x<0.\end{cases}$$

例 2.4.2 设 $M=\{(\xi_1,\xi_2)\in \mathbb{R}^2 \mid \xi_1^2+\xi_2^2\le 1\}$, 考虑 $\mathbb{R}^2$ 上的凸函数 f:

$$f(x)=\begin{cases}0, & x\in \operatorname{int} M,\\ \infty, & x\in \mathbb{R}^2\backslash M,\\ [0,\infty]\text{中任意值}, & |x|=1.\end{cases}$$

那么

$$(\operatorname{cl} f)(x)=\begin{cases}0, & x\in M,\\ +\infty, & x\in \mathbb{R}^2\backslash M.\end{cases}$$

上述例子表明, 通过对凸函数在某些点上重新赋值就能使之成为更正则的凸函数. 换句话说, 对一个给定的凸函数, 闭包运算不用对函数值作重大改变就能把它变为具有较好性质的闭的凸函数. 闭包运算在理论上和应用上的价值也正在于此.

下面我们进一步讨论凸函数 f 与其闭包 $\operatorname{cl} f$ 之间的关系.

引理 2.4.2 设 $f:\mathbb{R}^n\to[-\infty,\infty]$ 为凸函数. 那么

$$\operatorname{ri}(\operatorname{epi} f)=\{(x,\mu)\in \mathbb{R}^{n+1}\mid x\in \operatorname{ri}(\operatorname{dom} f)),\\ f(x)<\mu<\infty\}.$$

证明 实际上引理的结论可直接从定理 1.4.18 推出 (取 $m=1, M=\operatorname{epi} f$). 这里我们给出另一个证明. 用 M 表示上式右端的集合, 则显然有

$$\operatorname{ri}(\operatorname{epi} f)\subset M.$$

因此只须证明反包含关系 : $M\subset \operatorname{ri}(\operatorname{epi} f)$. 为此我们应用关于判别凸集相对内点的定理 1.4.11. 设 $x_0\in \operatorname{ri}(\operatorname{dom} f)$,

$f(x_0) < \lambda_0 < \infty$. 我们要证明 $(x_0, \lambda_0) \in \mathrm{ri}\,(\mathrm{epi}\, f)$. 任取 $(x, \lambda) \in \mathrm{epi}\, f$, 即 $f(x) \le \lambda < \infty$. 于是依据定理 1.4.11, 存在 $\mu_0 > 1$ 使

$$z = (1 - \mu_0)x + \mu_0 x_0 \in \mathrm{dom}\, f.$$

又根据定理 1.4.11, 如果能找到一个 $\mu > 1$, 使得

$$(1 - \mu)(x, \lambda) + \mu(x_0, \lambda_0) \in \mathrm{epi}\, f,$$

即

$$w = (1 - \mu)x + \mu x_0 \in \mathrm{dom}\, f,$$

$$f(w) \le (1 - \mu)\lambda + \mu\lambda_0,$$

则 $(x_0, \lambda_0) \in \mathrm{ri}\,(\mathrm{epi}\, f)$. 为此，对于 $0 < \nu < 1$ ，令

$$\begin{aligned} x(\nu) &= (1 - \nu)z - \nu x_0 \\ &= (1 - \nu)(1 - \mu_0)x + ((1 - \nu)\mu_0 + \nu)x_0 \\ &= (1 - \mu(\nu))x + \mu(\nu)x_0, \end{aligned}$$

其中 $\mu(\nu) = (1 - \nu)\mu_0 + \nu$. 显然 $1 < \mu(\nu) < \mu_0$. 再令 $\varepsilon = \lambda_0 - f(x_0)$, 则由 f 的凸性得

$$f(x(\nu)) \le (1 - \nu)f(z) + \nu f(x_0). \tag{2.4.2}$$

另一方面， x_0 可表示成 $x_0 = (1 - 1/\mu_0)x + z/\mu_0$, 于是同样由 f 的凸性得

$$f(z) \ge (1 - \mu_0)f(x) + \mu_0 f(x_0). \tag{2.4.3}$$

因此从式 (2.4.2) 和 (2.4.3) 可得

$$\begin{aligned} f(x(\nu)) \le{}& (1 - \mu(\nu))\lambda + \mu(\nu)\lambda_0 - \varepsilon\nu \\ &+ (1 - \nu)[f(z) + (\mu_0 - 1)\lambda - \lambda_0\mu_0]. \end{aligned}$$

今选择 $\nu \in (0,1)$ 使得

$$|(1/\nu - 1)[f(z) + (\mu_0 - 1)\lambda - \mu_0\lambda_0]| < \varepsilon,$$

并取 $\mu = \mu(\nu)$, 则 $\mu > 1$, 并且

$$f(1-\mu)x + \mu x_0) \leq (1-\mu)\lambda + \mu\lambda_0,$$

即

$$(1-\mu)(x,\mu) + \mu(x_0,\lambda_0) \in \operatorname{epi} f. \qquad \blacksquare$$

从闭包的定义直接可知: $\operatorname{cl} f \leq f$, 并且 $f_1 \leq f_2 \Longrightarrow \operatorname{cl} f_1 \leq \operatorname{cl} f_2$. 进一步我们有如下推论:

推论 2.4.3 设 f, g 为 $\mathbb{R}^n$ 上的两个真凸函数, 满足 $\operatorname{ri}(\operatorname{dom} f) = \operatorname{ri}(\operatorname{dom} g)$, 并且

$$f(x) = g(x), \quad \forall x \in \operatorname{ri}(\operatorname{dom} f)(= \operatorname{ri}(\operatorname{dom} g)).$$

那么 $\operatorname{cl} f = \operatorname{cl} g$.

证明 根据引理 2.4.2, 定理的假设意味着

$$\operatorname{ri}(\operatorname{epi} f) = \operatorname{ri}(\operatorname{epi} g).$$

于是由定理 1.4.3 得到

$$\operatorname{cl}(\operatorname{epi} f) = \operatorname{cl}(\operatorname{epi} g),$$

这正是说 $\operatorname{cl} f = \operatorname{cl} g$. $\blacksquare$

推论 2.4.4 设 f 是 $\mathbb{R}^n$ 上的真凸函数, $\lambda \in \mathbb{R}$ 和 $x \in \operatorname{dom} f$ 使得 $f(x) < \lambda$. 那么必定存在某个 $y \in \operatorname{ri}(\operatorname{dom} f)$ 使得 $f(y) < \lambda$.

证明 事实上, 假设 $f(x) < \lambda$ 意味着 $\mathbb{R}^{n+1}$ 中的开半空间 $\{(x,\mu) \mid x \in \mathbb{R}^n, \mu < \lambda\}$ 与上图 $\operatorname{epi} f$ 相交. 于是依据推论 1.4.7,

$$\{(x,\mu) \mid x \in \mathbb{R}^n, \mu < \lambda\} \cap \operatorname{ri}(\operatorname{epi} f) \neq \emptyset,$$

这正是说存在 $y \in \operatorname{ri}(\operatorname{dom} f)$ 使得 $f(y) < \lambda$. $\blacksquare$

推论 2.4.5 设 f 是 $\mathbb{R}^n$ 上的真凸函数, $M \subset \mathbb{R}^n$ 为一凸集使得 $\mathrm{ri}\, M \subset \mathrm{dom}\, f$. 假定对于 $\lambda \in \mathbb{R}$, 存在 $x \in \mathrm{cl}\, M$ 满足 $f(x) < \lambda$, 则必存在某个 $y \in \mathrm{ri}\, M$ 使得 $f(y) < \lambda$.

证明 定义凸函数 $g : \mathbb{R}^n \to (-\infty, \infty]$,

$$g(x) = \begin{cases} f(x), & x \in \mathrm{cl}\, M, \\ +\infty, & x \in \mathbb{R}^n \backslash \mathrm{cl}\, M. \end{cases}$$

显然 $\mathrm{ri}\,(\mathrm{dom}\, g) = \mathrm{ri}\, M$. 依假设 $g(x) < \lambda$, 从而由推论 2.4.4, 存在某个 $y \in \mathrm{ri}\,(\mathrm{dom}\, g) = \mathrm{ri}\, M$ 使得 $g(y) < \lambda$. 但 $\mathrm{ri}\, M \subset \mathrm{dom}\, f$, 因此 $g(y) = f(y)$. ▮

推论 2.4.6 设 f 为 $\mathbb{R}^n$ 上的真凸函数, 并且凸集 $M \subset \mathrm{dom}\, g$. 如果 $f(x) \geq \lambda$, $\forall\, x \in M$, 则 $f(y) \geq \lambda$, $\forall\, y \in \mathrm{cl}\, M$.

证明 事实上, 如果存在 $y \in \mathrm{cl}\, M$ 使 $f(y) < \lambda$, 则依推论 2.4.5, 必存在 $x \in \mathrm{ri}\, M$ 使 $f(x) < \lambda$, 导致矛盾. ▮

定理 2.4.7 $\mathbb{R}^n$ 上任意非真凸下半连续函数 f 不可能取有穷值.

证明 依据定理 2.2.11, 非真凸函数 f 在 $\mathrm{ri}\,(\mathrm{dom}\, f)$ 上取值 $-\infty$. 因此从下半连续函数性的定义可知:

$$x \in \mathrm{cl}\,(\mathrm{ri}\,(\mathrm{dom}\, f)) \Longrightarrow f(x) = -\infty.$$

但从定理 1.4.5 又可得到:

$$\mathrm{cl}\,(\mathrm{ri}\,(\mathrm{dom}\, f)) = \mathrm{cl}\,(\mathrm{dom}\, f) \supset \mathrm{dom}\, f.$$

由此可见定理结论成立. ▮

这个结果表明: $\mathbb{R}^n$ 上仅有的下半连续非真凸函数只有两个: 即恒等于 $+\infty$ 或 $-\infty$ 的两个常值函数.

下一个定理也许是关于闭包运算的最重要的结论.

定理 2.4.8 设 $f: \mathbb{R}^n \to (-\infty, \infty]$ 为真凸函数，那么 $\mathrm{cl}\, f$ 是闭的真凸函数；此外，$\mathrm{cl}\, f$ 与 f 至多在 $\mathrm{dom}\, f$ 的一些相对边界点上取不同值. 特别地由此可见 f 在 $\mathrm{ri}\,(\mathrm{dom}\, f)$ 中下半连续函数.

证明 仅需证明后一部分结论. 设 $x \in \mathrm{ri}\,(\mathrm{dom}\, f)$, 考虑 $\mathbb{R}^{n+1}$ 中的铅垂直线 $L = \{(x, \mu) \mid \mu \in \mathbb{R}\}$. 从引理 2.4.2 知 $L \cap \mathrm{ri}\,(\mathrm{epi}\, f) \neq \emptyset$. 于是根据定理 1.4.13,

$$L \cap \mathrm{cl}\,(\mathrm{epi}\, f) = \mathrm{cl}\,(L \cap \mathrm{epi}\, f) = L \cap \mathrm{epi}\, f.$$

(注意对于仿射集 L, $\mathrm{cl}\, L = L$) 这正是说

$$(\mathrm{cl}\, f)(x) = f(x).$$

最后，当 $x \notin \mathrm{cl}\,(\mathrm{dom}\, f)$ 时，依 $\mathrm{cl}\, f$ 的定义,

$$\mathrm{cl}\,(\mathrm{dom}\, f) \supset \mathrm{dom}\,(\mathrm{cl}\, f) \supset \mathrm{dom}\, f.$$

因此

$$\mathrm{cl}\, f(x) = f(x) = \infty.$$

∎

推论 2.4.9 如果 f 是 $\mathbb{R}^n$ 上的真凸函数, 则 $\mathrm{dom}\, f$ 与 $\mathrm{dom}\,(\mathrm{cl}\, f)$ 的差别至多是加上 $\mathrm{dom}\, f$ 的一些相对边界点. 特别 $\mathrm{dom}\,(\mathrm{cl}\, f)$ 与 $\mathrm{dom}\, f$ 有相同的闭包和相对内部以及相同的维数.

推论 2.4.10 如果 f 是 $\mathbb{R}^n$ 上的真凸函数，并且 $\mathrm{dom}\, f$ 是一仿射集，则 f 是闭的，即 f 是下半连续的.

上面的讨论表明，凸函数 f 除了 $\mathrm{dom}\, f$ 的某些边界点外总是下半连续的. 下一节中我们将看到，f 实际上相对于 $\mathrm{ri}\,(\mathrm{dom}\, f)$ 是连续的.

下一个定理为计算凸函数 f 的闭包 $\mathrm{cl}\, f$ 提供了更为简单的极限公式.

定理 2.4.11 设 f 是 $\mathbb{R}^n$ 上的真凸函数，并设 $x \in \mathrm{ri}\,(\mathrm{dom}\, f)$, 那么

$$(\mathrm{cl}\, f)(y) = \lim_{\lambda \uparrow 1} f((1-\lambda)x + y), \quad \forall\, y \in \mathbb{R}^n.$$

证明 由于 $\mathrm{cl}\, f$ 下半连续，并且 $\mathrm{cl}\, f \le f$, 我们有

$$(\mathrm{cl}\, f)(y) \le \liminf_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y).$$

于是我们只要证明

$$(\mathrm{cl}\, f)(y) \ge \limsup_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y).$$

假定 $\beta \in \mathbb{R}$ 满足 $\beta \ge (\mathrm{cl}\, f)(y)$, 并任取一实数 $\alpha > f(x)$. 于是,

$$(y, \beta) \in \mathrm{epi}\,(\mathrm{cl}\, f) = \mathrm{cl}\,(\mathrm{epi}\, f),$$

$$(x, \alpha) \in \mathrm{ri}\,(\mathrm{epi}\, f).$$

因此根据定理 1.4.1,

$$(1-\lambda)(x, \alpha) + \lambda(y, \beta) \in \mathrm{ri}\,(\mathrm{epi}\, f), \quad 0 \le \lambda < 1,$$

于是

$$f((1-\lambda)x + \lambda y) < (1-\lambda)\alpha + \lambda\beta, \quad 0 \le \lambda < 1.$$

由此得到

$$\limsup_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y) \le \limsup_{\lambda \uparrow 1}[(1-\lambda)\alpha + \lambda\beta] = \beta.$$

由于 $\beta \ge (\mathrm{cl}\, f)(y)$ 的任意性，定理得证. ∎

推论 2.4.12 设 f 是 $\mathbb{R}^n$ 上的闭真凸函数，则

$$f(y) = \lim_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y), \quad \forall\, x \in \mathrm{dom}\, f,\ y \in \mathbb{R}^n.$$

证明 设 $\varphi(\lambda) = f((1-\lambda)x + \lambda y), 0 \le \lambda \le 1$, 并用 ∞ 把 φ 延拓成全 $\mathbb{R}$ 上的凸函数. 显然 $\varphi(0) = f(x) < \infty, \varphi(1) = f(y)$. 从 f 的闭性容易推出 φ 的闭性, 并且 $\operatorname{dom}\varphi \subset [0,1]$. 如果 $y \in \operatorname{dom} f$, 则 $\operatorname{int}(\operatorname{dom}\varphi) = (0,1)$. 于是 $\lim\limits_{\lambda\uparrow 1}\varphi(\lambda) = \varphi(1)$. 如果 $y \notin \operatorname{dom} f$, 则当 $\operatorname{dom}\varphi$ 含内点时, 依据定理 2.4.11 有 $\lim\limits_{\lambda\uparrow 1}\varphi(\lambda) = (\operatorname{cl}\varphi)(1) = \varphi(1)$; 而当 $\operatorname{int}(\operatorname{dom}\varphi) = \emptyset$ 时, 极限 $\lim\limits_{\lambda\uparrow 1}\varphi(\lambda)$ 和 $\varphi(1)$ 显然都是 ∞. ∎

现在我们来讨论上一节中提出的凸函数表示成一族仿射函数的上包络的问题.

对于凸函数 $f : \mathbb{R}^n \to (-\infty, \infty]$, 我们用 $A(f)$ 表示不超过 f 的仿射函数全体, 换句话说, $a \in A(f)$ 意味着存在一个向量 $x_a^* \in \mathbb{R}^n$ 和一个实数 $\alpha_a \in \mathbb{R}$ 使

$$a(x) = \langle x_a^*, x\rangle + \alpha_a \le f(x), \quad \forall x \in \mathbb{R}^n.$$

定理 2.4.13 设 f 为 $\mathbb{R}^n$ 上的真凸函数, 则

$$f(x) = \sup_{a\in A(f)} a(x), \quad \forall x \in \operatorname{ri}(\operatorname{dom} f). \tag{2.4.4}$$

证明 只须证明对于任意 $x_0 \in \operatorname{ri}(\operatorname{dom} f)$, 存在一个 $a \in A(f)$ 使得 $a(x_0) = f(x_0)$. 事实上, 依据引理 2.4.2 有 $(x_0, f(x_0)) \notin \operatorname{ri}(\operatorname{epi} f)$. 于是由凸集分离定理 1.5.4, 存在非零 $(z,\gamma) \in \mathbb{R}^n \times \mathbb{R}$ 满足

$$\begin{cases} \langle z, x_0\rangle + \gamma f(x_0) \le \langle z, x\rangle + \gamma\alpha, \\ \quad \forall (x,\alpha) \in \operatorname{epi} f, \end{cases} \tag{2.4.5}$$

$$\begin{cases} \langle z, x_0\rangle + \gamma f(x_0) < \langle z, x\rangle + \gamma\alpha, \\ \quad \forall (x,\alpha) \in \operatorname{ri}(\operatorname{epi} f). \end{cases} \tag{2.4.6}$$

特别在式 (2.4.5) 中取 $x = x_0, \alpha < f(x_0)$ 可知 $\gamma \geq 0$. 同时 $\gamma = 0$ 也不可能，因为否则的话，从

$$(x_0, f(x_0) + \beta) \in \mathrm{ri}\,(\mathrm{epi}\, f), \quad \forall\, \beta > 0$$

推出式 (2.4.6) 不可能成立. 因此 $\gamma > 0$. 令

$$a(x) = \langle z, x\rangle/\gamma + < \langle z, x_0\rangle/\gamma + f(x_0), \qquad (2.4.7)$$

则由式 (2.4.5) 得到

$$a(x) \leq f(x), \quad \forall\, x \in \mathrm{dom}\, f.$$

从而 $a \in A(f)$, 并且由式 (2.4.7) 知 $a(x_0) = f(x_0)$. ∎

推论 2.4.14 设 $f: \mathbb{R}^n \to \mathbb{R}$ 为凸函数，则

$$f(x) = \sup_{a \in A(f)} a(x), \quad \forall\, x \in \mathbb{R}^n.$$

注意定理 2.4.13 中的结论不能推广到一切 $x \in \mathbb{R}^n$:

$$f(x) = \sup_{a \in A(f)} a(x), \quad \forall\, x \in \mathbb{R}^n. \qquad (2.4.8)$$

这是因为一族闭的真凸函数的上包络仍然是闭的. 因此，式 (2.4.8) 右端作为仿射函数的上包络一定是下半连续的. 但我们知道，凸函数未必都是下半连续的，例如 $\mathbb{R}$ 上的真凸函数:

$$f(x) \begin{cases} 0, & x \in (-1, 1), \\ 1, & x = \pm 1, \\ \infty, & |x| > 1, \end{cases}$$

在 $x = \pm 1$ 时就不是下半连续的. 从而式 (2.4.8) 不成立.

定理 2.4.15 设 $f: \mathbb{R}^n \to (-\infty, \infty]$ 为真凸函数，那么式 (2.4.8) 成立的充分必要条件是: 在全 $\mathbb{R}^n$ 上是下半连续的.

证明 只须证明充分性. 于是设 f 是下半连续真凸函数. 任取 $x_0 \in \operatorname{dom} f$ 和 $\varepsilon > 0$, 则

$$(x_0, f(x_0) - \varepsilon) \notin \operatorname{epi} f = \operatorname{cl}(\operatorname{epi} f).$$

于是由凸集强分离定理 1.5.6 和推论 1.5.7, 存在非零 $(z, \gamma) \in \mathbb{R}^n \times \mathbb{R}$ 使得

$$\langle z, x_0 \rangle + \gamma(f(x_0) - \varepsilon) < \langle z, x \rangle + \gamma\alpha, \quad \forall (x, \alpha) \in \operatorname{epi} f.$$

依照定理 2.4.13 中的推理可知 $\gamma > 0$. 令

$$a_\varepsilon(x) = -\langle z, x \rangle / \gamma + \langle z, x_0 \rangle / \gamma + (f(x_0) - \varepsilon),$$

则 $a_\varepsilon \in A(f)$, 并且

$$a_\varepsilon(x_0) = f(x_0).$$

由于 $\varepsilon > 0$ 的任意性, 这就证明了

$$f(x_0) = \sup_{a \in A(f)} a(x_0), \quad x_0 \in \operatorname{dom} f.$$

最后我们证明, 当 $x_1 \notin \operatorname{dom} f$, 即 $f(x_1) = \infty$ 时也有

$$\sup_{a \in A(f)} a(x_1) = \infty.$$

事实上, 由于 $x_1 \notin \operatorname{dom} f$, 故对任意 $\beta \in \mathbb{R}$, $(x_1, \beta) \notin \operatorname{epi} f$. 于是再次利用推论 1.5.7, 存在非零 $(z_1, \gamma_1) \in \mathbb{R}^n \times \mathbb{R}$ 和 $\mu \in \mathbb{R}, \delta > 0$, 使得

$$\begin{cases} \langle z_1, x_1 \rangle + \gamma_1 \beta = \mu - \delta < \mu \le \langle z_1, x \rangle + \gamma_1 \alpha, \\ \qquad \forall (x, \alpha) \in \operatorname{epi} f. \end{cases} \tag{2.4.9}$$

同前面一样, 我们仍然有 $\gamma_1 \ge 0$. 现在分别讨论两种情形:

(1) $\gamma_1 > 0$. 记 $a_\beta(x) = -\langle z_1, x \rangle / \gamma_1 + \langle z_1, x_1 \rangle / \gamma_1 + \beta$; 则 $a_\beta \in A(f)$, 并且 $a_\beta(x_1) = \beta$. 由于 $\beta > 0$ 的任意性,

$$\sup_{a\in A(f)} a(x_1)=\infty.$$

(2) $\gamma_1=0$. 从式 (2.4.9) 得到

$$0\geq\delta+\langle z_1,x_1-x\rangle,\quad \forall x\in\operatorname{dom}f. \tag{2.4.10}$$

另一方面，由于 $\operatorname{ri}(\operatorname{dom}f)\neq\emptyset$, 存在 $z_0\in\mathbb{R}^n$ 和 $\gamma_0\in\mathbb{R}$ 使得

$$f(x)\geq\langle z_0,x\rangle+\gamma_0,\quad \forall x\in\mathbb{R}^n, \tag{2.4.11}$$

即 f 下有界于一个仿射函数. 现在把式 (2.4.10) 两端乘以 n 并与 (2.4.11) 相加，得到

$$f(x)\geq n\delta+\langle nz_1,x_1-x\rangle+\langle z_0,x\rangle+\gamma_0,\quad \forall x\in\operatorname{dom}f.$$

令

$$a_n(x)=\langle z_0-nz_1,x\rangle+n\delta+\gamma_0+\langle nz_1,x_1\rangle,$$

则 $a_n\in A(f)$, 并且

$$a_n(x_1)=n\delta+\gamma_0+\langle z_0,x_1\rangle.$$

由于 n 可以任意大，所以我们得到

$$\sup_{a\in A(f)} a(x_1)=\infty.$$ ∎

在结束本节讨论时，我们再对 f 的函数回收方向作一些说明. 定理 2.3.11 指出，对于 $\mathbb{R}^n$ 上的真凸函数 f, 为了向量 $y\in\mathbb{R}^n$ 属于 f 的回收维，必须且只须对于任意 $x\in\mathbb{R}^n$, $f(x+\lambda y)$ 作为 $\lambda\in\mathbb{R}$ 的函数是不递增的. 当 f 还是闭的时候，上述结论可以加强. 我们有

定理 2.4.16 设 f 为 $\mathbb{R}^n$ 上的真凸函数，$y\in\mathbb{R}^n$. 对于给定的 $x\in R^n$, 如果有

$$\liminf_{\lambda\to+\infty} f(x+\lambda y) < +\infty, \tag{2.4.12}$$

则 $f(x+\lambda y)$ 作为 $\lambda \in I\!R$ 的函数是不递增的. 当 f 还是闭的时候, 只要 (2.4.12 对于某个 $x \in \operatorname{dom} f$ 成立, $f(x+\lambda y)$ 对于任意 $x \in I\!R^n$ 作为 $\lambda \in I\!R$ 的函数是不递增的, 从而 y 是 f 的一个回收方向.

证明 如果式 (2.4.12) 成立, 则存在 $\alpha \in I\!R$ 使得

$$\liminf_{\lambda\to+\infty} f(x+\lambda y) < \alpha < +\infty, \tag{2.4.13}$$

令 $g(\lambda) = f(x+\lambda y)$, 则 g 是 $I\!R$ 上的真凸函数, 并且从式 (2.4.13) 可知存在序列 $\{(\lambda_k, \alpha) \mid k \geq 1\} \subset \operatorname{epi} g$, 其中 $\lambda_k \to +\infty (k \to \infty)$. 序列 $\{(\lambda_k, \alpha) \mid k \geq 1\}$ 的凸包是沿着向量 $(1,0) \in I\!R^2$ 方向的一条半直线, 并且包含在闭凸集 $\operatorname{epi}(\operatorname{cl} g)$ 中. 因此向量 (1,0) 是 $\operatorname{epi}(\operatorname{cl} g)$ 的一个回收方向, 从而 $\operatorname{cl} g$ 是 $I\!R$ 上的一个非增函数. $\operatorname{cl} g$ 的有效定义域 $\operatorname{dom}(\operatorname{cl} g)$ 必定是 $I\!R$ 上的一个右无界的区间. 依据定理 2.4.8, $\operatorname{cl} g$ 仅仅在其有效定义域的边界点上可能跟 g 取不同值, 因此 g 本身, 即 $f(x+\lambda y)$ 是 $I\!R$ 上的非增函数.

如果 f 还是闭的, 则根据定理 2.3.10 中的公式 (2.3.4), 当存在一点 $x \in \operatorname{dom} f$ 使得 $f(x+\lambda y)$ 为 λ 的非增函数时有 $(f0^+)(y) \leq 0$. ∎

推论 2.4.17 设 f 是 $I\!R^n$ 上的真凸函数, $y \in I\!R^n$. 那么为了对于每一个 $x \in I\!R^n$, $f(x+\lambda y)$ 是 $\lambda \in I\!R$ 的常值函数, 必须且只须 $(f0^+)(y) \leq 0$ 和 $(f0^+)(-y) \leq 0$. 在 f 还是闭的情况下, 只要存在一个 x 和某个实数 α 使得

$$f(x+\lambda y) \leq \alpha, \quad \forall \lambda \in I\!R,$$

则上述条件就满足.

集合 $\{y \in \mathbb{R}^n \mid (f0^+)(y) \leq 0, (f0^+)(-y) \leq 0\}$ 是包含在 f 的回收锥中的一个最大的子空间，我们把它称作 f 的常值空间 (constancy space). f 的常值空间中的向量叫做 f 的常值方向.

定理 2.4.18 设 f 是 $\mathbb{R}^n$ 上的闭真凸函数. 那么 f 的所有非空的水平集 $S_\alpha(f) = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}, \alpha \in \mathbb{R}$, 都有相同的回收锥和相同的直线状空间，也就是 f 的回收锥和常值空间.

证明 只须注意依据定理 2.4.16, y 属于 $S_\alpha(f)$ 的回收锥当且仅当

$$f(x) \leq \alpha,\ \lambda \geq 0 \Longrightarrow f(x + \lambda y) \leq \alpha. \qquad \blacksquare$$

推论 2.4.19 设 f 是 $\mathbb{R}^n$ 上的真凸函数，如果水平集 $S_\alpha(f) = \{x \mid f(x) \leq \alpha\}$ 对于某一个 α 是非空且有界的，它对于一切 α 均是有界的.

定理 2.4.20 设 f 是 $\mathbb{R}^n$ 上的真凸函数， $M \subset \mathbb{R}^n$ 为一非空闭凸集. 假定 $g = f + \delta(\cdot|M)$ 仍是真凸的，那么 y 是 g 的回收方向当且仅当 y 是 f 和 M 的公共回收方向.

证明 留作练习. $\blacksquare$

定理 2.4.21 设 f 是 $\mathbb{R}^n$ 上的真凸函数, 并设 $\alpha \in \mathbb{R}$, $\alpha > \inf\{f(x) \mid x \in \mathbb{R}^n\}$. 那么水平集 $\{x \mid f(x) \leq \alpha\}$ 和 $\{x \mid f(x) < \alpha\}$ 有相同的闭包和相同的相对内部, 也即分别是

$$\{x \mid (\mathrm{cl}\, f)(x) \leq \alpha\} \text{ 和 } \{x \in \mathrm{ri}\,(\mathrm{dom}\, f) \mid f(x) < \alpha\}.$$

证明 设 $M = \{(x, \alpha) \mid x \in \mathbb{R}^n\}$ 为 $\mathbb{R}^{n+1}$ 的水平超平面. 根据引理 2.4.2, $M \cap \mathrm{ri}\,(\mathrm{epi}\, f) \neq \emptyset$. 注意

$$M \cap \mathrm{epi}\, f = \{(x, \alpha) \mid f(x) \leq \alpha\}.$$

依据定理 1.4.13, $M\cap\operatorname{epi}f$ 的闭包和相对内部分别为 $M\cap\operatorname{cl}(\operatorname{epi}f)$ 和 $M\cap\operatorname{ri}(\operatorname{epi}f)$. 但 $\operatorname{cl}(\operatorname{epi}f)=\operatorname{epi}(\operatorname{cl}f)$, 因此,

$$\operatorname{cl}\{x\mid f(x)\le\alpha\}=\{x\mid(\operatorname{cl}f)(x)\le\alpha\},\tag{2.4.14}$$

$$\operatorname{ri}\{x\mid f(x)\le\alpha\}=\{x\in\operatorname{ri}(\operatorname{dom}f)\mid f(x)<\alpha\}.\tag{2.4.15}$$

从式 (2.4.15) 得到

$$\operatorname{ri}\{x\mid f(x)\le\alpha\}\subset\{x\mid f(x)<\alpha\}\subset\{x\mid f(x)\le\alpha\}.$$

因此 $\{x\mid f(x)<\alpha\}$ 与 $\{x\mid f(x)\le\alpha\}$ 具有相同的闭包和相同的相对内部. ∎

习　题　2.4

2.4.1 设 $f:\mathbb{R}^n\to(-\infty,\infty]$ 为凸函数. 如果存在一点 $x_0\in\operatorname{ri}(\operatorname{dom}f)$ 使得

$$f(x_0)=\sup\{f(x)\bigm| x\in\mathbb{R}^n\}.$$

求证 f 为常值函数.

2.4.2 设 M 是 $\mathbb{R}^n$ 的非空闭凸子集. 试证: 示性函数 $\delta(x|M)$ 是 $\mathbb{R}^n$ 上的下半连续真凸函数.

2.4.3 研究下列函数的下半连续性:

$$f(x)=\begin{cases}1/x, & \forall\, x>0,\\ +\infty, & \forall\, x\le 0;\end{cases}\qquad g(x)=\begin{cases}x^2, & \forall\,|x|<1,\\ +\infty, & \forall\,|x|\ge 1.\end{cases}$$

2.4.4 设 M 是 $\mathbb{R}^n$ 的非空闭凸集. 试证: $\delta(\cdot|M)$ 和 M 具有相同的回收方向.

2.4.5 设 f 为 $\mathbb{R}^n$ 上的真凸函数, 试证: y 为 f 的回收方向, 即 $(f0^+)(y)\le 0$, 当且仅当 $(y,0)\in 0^+(\operatorname{epi}f)$, 或者当且仅当对于任意满足 $f(x)\le\nu$ 的 $(x,\nu)\in\mathbb{R}^{n+1}$ 有

$$f(x+\lambda y)\le\nu,\quad \forall\,\lambda\ge 0.$$

而当 f 还是闭的时候，则 y 是 f 的回收方向当且仅当对于某个 x，从 $(x,\nu)\in \operatorname{epi} f$ 出发的"水平"方向的半直线 $\{(x+\lambda y,\nu)\mid \lambda\geq 0\}$ 都包含在 $\operatorname{epi} f$ 中.

2.4.6 完成定理 2.4.20 的证明.

2.5 凸函数的连续性

上一节我们指出 $\mathbb{R}^n$ 上的真凸函数 f 在其有效定义域的相对内部 $\operatorname{ri}(\operatorname{dom} f)$ 中是下半连续的，并指出适当改变其在相对边界上的值就成为 $\mathbb{R}^n$ 上的下半连续函数. 现在我们进一步证明，$\mathbb{R}^n$ 上的真凸函数 f 相对 $\operatorname{ri}(\operatorname{dom} f)$ 还是连续的.

$\mathbb{R}^n$ 上的函数称为相对于 $\mathbb{R}^n$ 的某个子集 M 是连续的，是指 f 限制在 M 上是一个连续函数；换句话说，f 相对于 M 连续，是指对任意 $x_0\in M$，当 x 沿着 M 趋于 x_0 时 $f(x)$ 收敛于 $f(x_0)$，即

$$\lim_{\substack{x\to x_0\\ x\in M}} f(x)=f(x_0).$$

定理 2.5.1 设 $f:\mathbb{R}^n\to(-\infty,+\infty]$ 为真凸函数，那么 f 相对其有效定义域 $\operatorname{dom} f$ 的相对内部 $\operatorname{ri}(\operatorname{dom} f)$ 是连续的.

证明 我们在定理 2.4.4 中已经指出 f 在 $\operatorname{ri}(\operatorname{dom} f)$ 中的下半连续性. 为证 f 相对于 $\operatorname{ri}(\operatorname{dom} f)$ 的连续性，只须证明 f 相对于 $\operatorname{ri}(\operatorname{dom} f)$ 的上半连续性. 于是任取 $x_0\in\operatorname{ri}(\operatorname{dom} f)$，并设

$$x_k\in\operatorname{ri}(\operatorname{dom} f),\ \lim_{k\to\infty}x_k=x_0,\ \lim_{k\to\infty}f(x_k)=\mu.$$

今证 $\mu\leq f(x_0)$. 事实上，否则，$f(x_0)<\mu$，于是由引理

2.4.2, $(x_0,\mu)\in \mathrm{ri}\,(\mathrm{epi}\,f)$. 但是

$$(x_k, f(x_k))\in \mathrm{epi}\,f,\quad \lim_{k\to\infty}(x_k,f(x_k))=(x_0,\mu),$$

故当 k 充分大后， $(x_k,f(x_k))\in \mathrm{ri}\,(\mathrm{epi}\,f)$. 但这是不可能的. ∎

推论 2.5.2 凸函数 $f:\mathbb{R}^n\to\mathbb{R}$ 是 $\mathbb{R}^n$ 上的连续函数.

上述定理的重要性在于函数的凸性隐含了函数一定程度的连续性. 例如，设 I 为某一指标集，对于每一 $i\in I, f_i:\mathbb{R}^n\to\mathbb{R}$ 为凸函数，且

$$f_i(x)\le C(x),\quad \forall\, i\in I, \forall\, x\in\mathbb{R}^n$$

其中常数 $C(x)$ 仅与 $x\in\mathbb{R}^n$ 有关，而与 $i\in I$ 无关. 那么该函数族的上包络

$$h(x)\triangleq \sup\{f_i(x)\mid i\in I\},\quad \forall\, x\in\mathbb{R}^n$$

在 $\mathbb{R}^n$ 上也是连续的.

定义在集合 $M\subset\mathbb{R}^n$ 上的实值函数 f 称之为相对于 M 是 Lipschitz 连续的，是指存在一个常数 $C\ge 0$ 使得

$$|f(x)-f(y)|\le C\|x-y\|,\quad \forall\, x,y\in M.$$

特别地， Lipschitz 连续性蕴涵着一致连续性.

定理 2.5.3 设 $f:\mathbb{R}^n\to(-\infty,+\infty]$ 为真凸函数，M 为 $\mathrm{ri}\,(\mathrm{dom}\,f)$ 的任意闭有界子集. 那么 f 相对于 M 是 Lipschitz 连续的.

证明 用 $\overline{B}_M(0,\varepsilon)$ 表示 $\mathrm{aff}\,M$ 中以 0 为中心、ε 为半径的闭球，即 $\overline{B}_M(0,\varepsilon)\triangleq\{x\in \mathrm{aff}\,M\mid \|x\|\le\varepsilon\}$. 由于 M

是 $\text{ri}(\text{dom} f)$ 中的有界闭集，故必存在 $\varepsilon > 0$, 使得

$$M + \overline{B}_M(0, \varepsilon) \subset \text{ri}(\text{dom} f).$$

依据定理 2.5.1, f 在 $M + \overline{B}_M(0, \varepsilon)$ 上相对 $\text{ri}(\text{dom} f)$ 是连续的. 但显然 $M + \overline{B}_M(0, \varepsilon)$ 是有界闭集，故 f 在 $M + \overline{B}_M(0, \varepsilon)$ 上有界，即存在实数 α_1 和 α_2, 使得

$$\alpha_1 \leq f(x) \leq \alpha_2, \quad \forall x \in M + \overline{B}_M(0, \varepsilon).$$

现在任取两不同点 $x, y \in M$, 并令

$$z = y + (\varepsilon / \|x - y\|)(y - x).$$

那么 $z \in M + \overline{B}_M(0, \varepsilon)$, 并且

$$y = (1 - \lambda)x + \lambda z, \quad \lambda \triangleq \|x - y\| / (\varepsilon + \|x - y\|).$$

显然 $0 < \lambda < 1$. 于是，由 f 的凸性得到

$$f(y) \leq (1 - \lambda) f(x) + \lambda f(z) = f(x) + \lambda(f(z) - f(x)).$$

由此得到

$$\begin{aligned} f(y) - f(x) \leq \lambda(\alpha_2 - \alpha_1) &= \frac{\|x - y\|}{\varepsilon + \|x - y\|}(\alpha_2 - \alpha_1) \\ &\leq \alpha \|y - x\|, \quad \alpha \triangleq (\alpha_2 - \alpha_1)/\varepsilon. \end{aligned}$$

上式对一切 $x, y \in M$ 成立，故这就证明了 f 相对于 M 的 Lipschitz 连续性. ▮

对于 $\mathbb{R}^n$ 的子集 M 上的函数族 $\{f_i \mid i \in I\}$ (I 为某个指标集), $\{f_i \mid i \in I\}$ 称为相对 M 等度 Lipschitz 连续，是指存在实数 $\alpha \geq 0$, 使得

$$|f_i(x) - f_i(y)| \leq \alpha \|x - y\|, \quad \forall x, y \in M, \forall i \in I.$$

特别地, 相对于 M 等度 Lipschitz 连续的函数族 $\{f_i \mid i \in I\}$ 必是相对 M 一致等度连续的，即对任意给定的 $\varepsilon > 0$, 存在 $\delta = \delta(\varepsilon) > 0$, 使得

$$x, y \in M,\ \|x - y\| \le \delta \Longrightarrow |f_i(x) - f_i(y)| \le \varepsilon,\ \forall i \in I.$$

函数族 $\{f_i \mid i \in I\}$ 称为在 M 上点点有界，是指对于每一 $x \in M$, 存在仅与 x 有关的两个实数 $C_1(x)$ 和 $C_2(x)$, 使得

$$C_1(x) \le f_i(x) \le C_2(x), \quad \forall i \in I.$$

如果上述常数 $C_1(x), C_2(x)$ 与 x 也无关，则称 $\{f_i \mid i \in I\}$ 在 M 上一致有界.

定理 2.5.4 设 $M \subset I\!R^n$ 为相对开凸集，又设 $\{f_i \mid i \in I\}$ 为 M 上一族实值凸函数 (I 为某指标集). 假定 $\{f_i \mid i \in I\}$ 在 M 上点点有界，并设 $S \subset M$ 为闭有界子集. 那么 $\{f_i \mid i \in I\}$ 在 S 上一致有界，并且相对 S 等度 Lipschitz 连续.

证明 令

$$f(x) = \sup\{f_i(x) \mid i \in I\}, \quad x \in M.$$

由 $\{f_i \mid i \in I\}$ 的点点有界性假设， f 是 M 上的有穷值凸函数，并且 $\operatorname{dom} f = M$ 为相对开集. 因此由定理 2.5.1, f 相当于 M 是连续的. 特别地， f 在 M 的闭有界子集 S 上一致上有界，即存在常数 $C_2 \in I\!R$ 使得

$$f(x) \le C_2, \quad \forall x \in S.$$

于是从 f 的定义得到

$$f_i(x) \le C_2, \quad \forall x \in S, \quad \forall i \in I.$$

为证明 $\{f_i \mid i \in I\}$ 在 S 上一致下有界，只要构造一个实值

连续函数 $g: M \to \mathbb{R}$, 使得

$$f_i(x) \geq g(x), \quad \forall x \in M, \quad \forall i \in I. \tag{2.5.1}$$

任意取定一个点 $x_0 \in M$, 由假设得到

$$-\infty < \alpha_1 = \inf\{f_i(x_0) \mid i \in I\}.$$

选择 $\varepsilon > 0$ 充分小使得

$$\overline{B}_M(x_0, \varepsilon) \subset M,$$

这里 $\overline{B}_M(x_0,\varepsilon) = \{x \in \operatorname{aff} M \mid \|x - x_0\| \leq \varepsilon\}$. 又设 α_2 为 f 在闭球 $\overline{B}_M(x_0,\varepsilon)$ 上的一个正上界. 对于任意 $x \in M, x \neq x_0$, 令

$$z \triangleq x_0 + \frac{\varepsilon}{\|x - x_0\|}(x_0 - x),$$

则

$$x_0 = (1-\lambda)z + \lambda x, \quad \lambda \triangleq \varepsilon/(\varepsilon + \|x - x_0\|).$$

由于 $z \in \overline{B}_M(x_0,\varepsilon) \subset M, 0 < \lambda < 1$, 由 f 的凸性得到

$$\begin{aligned}\alpha_1 \leq f_i(x_0) &\leq (1-\lambda)f_i(z) + \lambda f_i(x) \\ &\leq (1-\lambda)\alpha_2 + \lambda f_i(x) < \alpha_2 + \lambda f_i(x).\end{aligned}$$

于是

$$f_i(x) \geq (\alpha_1 - \alpha_2)/\lambda = (\alpha_1 - \alpha_2)(\varepsilon + \|x - x_0\|)/\varepsilon,$$
$$\forall x \in M, \ \forall i \in I.$$

若令 $g(x) \triangleq (\alpha_1 - \alpha_2)(\varepsilon + \|x - x_0\|)/\varepsilon$, 则 $g: M \to \mathbb{R}$ 是满足式 (2.1.5) 的连续函数. ∎

定理 2.5.5 设 $M \in R^n$ 为相对开凸集, $T \subset \mathbb{R}^m$ 为某一集合, 又设 $f: M \times T \to \mathbb{R}$. 假定对任意固定的 $t \in T, f(\cdot, t)$ 是 M 上的凸函数; 而对任意固定的 $x \in M, f(x, \cdot)$ 是 T 上的连续函数. 那么 f 是 $M \times T$ 上的连续函数.

证明 任取 $(x_0,t_0)\in M\times T$. 对于某个 $\sigma>0$, 令 $T_\sigma \triangleq \{t\in T \mid \|t-t_0\|\le\sigma\}$. T_σ 为 T 中的紧子集. 由假设, 对任意 $x\in M$, 函数 $f(x,\cdot)$ 在 T_σ 中有界. 因此, $\{f(\cdot,T)\mid t\in T_\sigma\}$ 为 M 上的一族点点有界的有穷值凸函数. 由定理 2.5.4, 这一族函数在 M 的任意闭有界集上等度 Lipschitz 连续. 于是给定 $\varepsilon>0$, 存在 $\delta_1>0$ 使得对于任意 $t\in T_\sigma$,

$$x\in M, \|x-x_0\|\le\delta_1 \Longrightarrow |f(x,t)-f(x_0,t)|\le\varepsilon/4,$$

设 $x_1\in M$ 满足 $\|x_1-x_0\|\le\delta_1$. 由于 $f(x_1,\cdot)$ 在 t_0 处连续, 存在 $\delta_2, 0<\delta_2\le\sigma$, 使得

$$t\in T,\ \|t-t_0\|\le\delta_2 \Longrightarrow |f(x_1,t)-f(x_1,t_0)|\le\varepsilon/4.$$

这样, 当 $(x,t)\in M\times T$ 满足 $\|x-x_0\|\le\delta_1, \|t-t_0\|\le\delta_2$ 时, 有

$$\begin{aligned}
&|f(x,t)-f(x_0,t_0)|\\
\le\ &|f(x,t)-f(x_0,t)|+|f(x_0,t)-f(x_1,t)|\\
&+|f(x_1,t)-f(x_1,t_0)|+|f(x_1,t_0)-f(x_0,t_0)|\\
\le\ &\varepsilon/4+\varepsilon/4+\varepsilon/4+\varepsilon/4=\varepsilon.
\end{aligned}$$

这就证明了 f 在 (x_0,t_0) 处的连续性. ∎

定理 2.5.6 设 $f_k: M\to\mathbb{R}$ 为凸函数, $k=1,2,\cdots$, $M\subset\mathbb{R}^n$ 为相对开集. 假定序列 $(f_k(x))$ 在 M 的一个稠子集 M' 上点点收敛, 且极限均有穷. 那么序列 $(f_k(x))$ 的极限对一切 $x\in M$ 都存在, 且极限函数

$$f(x)\triangleq\lim_{k\to\infty}f_k(x), \quad \forall x\in M$$

是 M 的有穷值凸函数. 此外，序列 (f_k) 在 M 的每一有界闭子集上一致收敛于 f.

证明 为记号简单起见，同时也不失一般性，假定 M 为 $\mathbb{R}^n$ 的开集. 设 S 是 M 的任一有界闭子集，于是存在 M 的另一个有界闭子集 S', 使得

$$S \subset \operatorname{int} S' \subset S' \subset M.$$

(见习题 1.2.15) 根据定理 2.5.5, $\{f_i\}$ 在 S' 上是等度 Lipschtz 连续的，即存在常数 $C > 0$, 使得

$$|f_j(x) - f_j(y) \leq C\|x - y\|, \quad \forall\, x, y \in S', \forall\, j \geq 1.$$

由于 S' 是紧集，并且 M' 在 S' 中稠，故对于任意给定的 $\varepsilon > 0$ 存在 $S' \cap M'$ 的一有限子集 S_0' 使得对于每一 $x \in S'$, 至少能找到一点 $z \in S_0'$ 满足 $\|x - z\| < \varepsilon/3C$. 既然 $\{f_j\}$ 在 S_0' 上点点收敛，而且 S_0' 是有限点集，故存在一自然数 j_0, 使得

$$|f_j(z) - f_i(z)| < \varepsilon/3, \quad \forall\, i, j \geq j_0, \forall\, z \in S_0'.$$

现在给定 $x \in S$, 取 $z \in S_0'$ 使得 $\|x - z\| < \varepsilon/3C$. 于是对于 $i, j \geq j_0$, 我们有

$$\begin{aligned}
&\|f_i(x) - f_j(x)| \\
&\leq |f_i(x) - f_i(z)| + |f_i(z) - f_j(z)| + |f_j(z) - f_j(x)| \\
&\leq C\|x - z\| + \varepsilon/3 + C\|x - z\| < \varepsilon.
\end{aligned}$$

这样，我们证明了 $\{f_j(x)\}$ 是一收敛序列. 记

$$f(x) = \lim_{j \to \infty} f_j(x), \quad \forall\, x \in S,$$

则 $f(x)$ 为有穷数, 并且对于给定的 $\varepsilon > 0$, 存在 $j_0 \geq 1$ 使得

$$|f_j(x) - f(x)| \leq \varepsilon, \quad \forall\, x \in S, \forall\, j \geq j_0.$$

这表明 (f_j) 在 S 上一致收敛于 f. 由于 S 是 M 的任意有界闭子集，f 定义在整个 M 上，并且处处取有穷值. 最后 f 的凸性可从每个 f_j 的凸性：

$$f_j((1-\lambda)x+\lambda y) \le (1-\lambda)f_j(x)+\lambda f_j(y),$$

$$\forall j \le 1, \quad \forall x,y \in M.$$

取极限得到. ∎

推论 2.5.7 设 M 为 $\mathbb{R}^n$ 的相对开凸集. 设 f 是 M 上的有穷值凸函数，而且 M 上有穷值凸函数序列 (f_j) 满足

$$\limsup_{j\to\infty} f_j(x) \le f(x), \quad \forall x \in M.$$

那么对于 M 的每一个闭有界子集 S 和 $\varepsilon>0$, 存在指标 j_0, 使得

$$f_j(x) \le f(x)+\varepsilon, \quad \forall j \ge j_0, \forall x \in S.$$

证明 令 $g_j(x)=\max\{f_j(x),f(x)\}$. 不难看出有穷值凸函数序列 (g_j) 在 M 上点点收敛于 f, 因此它在 S 上一致收敛于 f. ∎

习 题 2.5

2.5.1 设 T 为任一集合，并设 $f: \mathbb{R}^n \times T \to \mathbb{R}$. 假定对于每一固定的 $t \in T, f(x,t)$ 是 $x \in \mathbb{R}^n$ 的凸函数，而对于每一固定的 $x \in \mathbb{R}^n, f(x,t)$ 作为 t 的函数是上有界的，那么

$$g(x)=\sup\{f(x,t) \mid t \in T\}, \quad x \in \mathbb{R}^n$$

是 $\mathbb{R}^n$ 上的连续函数.

2.5.2 设 $f: \mathbb{R}^n \to \mathbb{R}$ 为凸函数，$M \subset \mathbb{R}^n$ 为一非空凸集. 令 $g(x)=\delta(x|-M)$. 试证：

$$h(x)=(f\Box g)(x)=\inf\{f(x-z)+\delta(z|-M) \mid z \in \mathbb{R}^n\}$$

是 $\mathbb{R}^n$ 上的连续函数.

2.5.3 在 $\mathbb{R}^2$ 上定义函数

$$f(\xi_1,\xi_2)=\begin{cases}\xi_2^2/2\xi_1, & \xi_1>0,\\ 0, & \xi_1=\xi_2=0,\\ +\infty, & \text{其余处}.\end{cases}$$

分析该函数在其有效定义域相对边界上的连续性.

2.5.4 设 f 是 $\mathbb{R}^n$ 上有穷值凸函数. 试证: 为了 f 在整个 $\mathbb{R}^n$ 上一致连续, 必须且只须 f 的回收函数 $f0^+$ 是处处有穷的. 这时 f 实际上是 $\mathbb{R}^n$ 上的 Lipschitz 连续函数.

2.6 某些闭性判据

凸集和凸函数有许多运算, 这些运算对凸集和凸函数的闭性会带来什么影响? 例如, 我们已经知道, 对于给定的凸集 $M\subset\mathbb{R}^n$ 和线性变换 $A:\mathbb{R}^n\to\mathbb{R}^m$, 我们有 $\mathrm{ri}\,(AM)=A(\mathrm{ri}\,M)$. 但对于闭包运算一般只有 $\mathrm{cl}\,(Am)\subset A(\mathrm{cl}\,M)$. 在什么条件下才会有 $\mathrm{cl}\,(AM)=A(\mathrm{cl}\,M)$ 呢? 在凸函数运算中也会遇到类似的问题, 需要我们仔细地去考虑.

我们在 1.8 节中已经定义了凸集 M 的回收锥 0^+M 的概念, 它对于解决凸集的无界性问题起了很大的作用. 对于非空凸集 $M\subset\mathbb{R}^n$, 我们知道集合 $(-0^+M)\cap 0^+M$ 是 M 的直线状空间 (lineality space), 并且有

$$(-0^+M)\cap 0^+M=\{y\in\mathbb{R}^n\mid M+y=M\}$$

对于 $\mathbb{R}^n$ 中的凸集 M 由集合 $\{(x,1)\mid x\in M\}$ 在 $\mathbb{R}^{n+1}$ 中生成的凸锥记作 K, 则显然

$$K=\{(x,\lambda)\mid\lambda>0,x\in\lambda M\}.$$

引理 2.6.1 设 $M \subset \mathbb{R}^n$ 为一闭凸集，K 为由 $\{(x,1) \mid x \in M\}$ 在 $\mathbb{R}^{n+1}$ 中生成的凸锥．那么

$$\mathrm{cl}\, K = K \cup \{(x,0) \mid x \in 0^+M\}.$$

证明 记 $K' = K \cup \{(x,0) \mid x \in 0^+M\}$. 如果 $x \in 0^+M$, 今证 $(x,0) \in \mathrm{cl}\, K$. 事实上，依定理 1.8.1, 存在 $x_k \in M, \lambda_k \downarrow 0 (k \to \infty)$ 使得 $\lambda_k x_k \to x(k \to \infty)$. 但 $(\lambda_k x_k, \lambda_k) \in K$, 因此 $(x,0) \in \mathrm{cl}\, K$. 剩下证明 $\mathrm{cl}\, K \subset K'$. 设 $(y,\mu) \in \mathrm{cl}\, K$, 则存在 $\lambda_k > 0, x_k \in \lambda_k M$ 使得 $(x_k, \lambda_k) \to (y,\mu)$. 如果 $\mu > 0$, 则 $(y,\mu) \in K$; 如果 $\mu = 0$, 则根据定理 1.8.1, $y \in 0^+M$. ▮

定理 2.6.2 设 $M \subset \mathbb{R}^n$ 为一非空凸集，$A : \mathbb{R}^n \to \mathbb{R}^m$ 为一线性变换．记 $\mathcal{N}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$ (A 的零空间). 假定

$$0^+(\mathrm{cl}\, M) \cap \mathcal{N}(A) \subset (-0^+(\mathrm{cl}\, M)) \cap 0^+(\mathrm{cl}\, M).$$

那么

$$\mathrm{cl}\,(AM) = A(\mathrm{cl}\, M), \quad 0^+A(\mathrm{cl}\, M) = A(0^+(\mathrm{cl}\, M)).$$

特别是，若 M 还是闭的，并且 $(0^+M) \cap \mathcal{N}(A) = \{0\}$, 则 AM 是闭集．

证明 只须证明 $\mathrm{cl}\,(AM) \subset A(\mathrm{cl}\, M)$, 即对于任意 $y \in \mathrm{cl}\,(AM)$, 存在 $x \in \mathrm{cl}\, M$ 使得 $y = Ax$. 为此令

$$L = (-0^+(\mathrm{cl}\, M)) \cap 0^+(\mathrm{cl}\, M) \cap \mathcal{N}(A),$$

则 L 是 $\mathbb{R}^n$ 中的一个子空间，并且根据有关 $0^+(\mathrm{cl}\, M)$ 的假设，

$$L = 0^+(\mathrm{cl}\, M) \cap \mathcal{N}(A).$$

由于 $\mathrm{cl}\, M$ 有直和分解 $\mathrm{cl}\, M = L + L^{\perp} \cap (\mathrm{cl}\, M))$, 我们有

$$A(\mathrm{cl}\, M) = A(L^{\perp} \cap (\mathrm{cl}\, M)).$$

现在设 $y \in \mathrm{cl}\,(AM)$. 对于任意 $\varepsilon > 0$, 如果记 $D_\varepsilon(A) = \{x \in \mathbb{R}^n \mid \|y - Ax\| \leq \varepsilon\}$, 则

$$M_\varepsilon = L^{\perp} \cap (\mathrm{cl}\, M) \cap D_\varepsilon(A) \neq \emptyset.$$

但根据推论 1.8.2,

$$\begin{aligned} 0^+ M_\varepsilon &= 0^+ L^{\perp} \cap 0^+(\mathrm{cl}\, M) \cap 0^+ D_\varepsilon(A) \\ &= L^{\perp} \cap 0^+(\mathrm{cl}\, M) \cap \mathcal{N}(A) = L^{\perp} \cap L = \{0\}, \end{aligned}$$

因此从定理 1.8.3 知 M_ε 是有界闭集. 此外 M_ε 显然随 ε 递减而递减: $0 < \varepsilon_1 < \varepsilon_2 \Longrightarrow M_{\varepsilon_1} \subset M_{\varepsilon_2}$. 这样, 根据数学分析中熟知的有关集合套的结果, 得到

$$\emptyset \neq \cap\{M_\varepsilon \mid \varepsilon > 0\} = L^{\perp} \cap (\mathrm{cl}\, M) \cap A^{-1}y,$$

从而存在 $x \in \mathrm{cl}\, M$, 使得 $Ax = y$.

剩下证明当 M 还是闭凸集时 $A(0^+ M) = 0^+(AM)$. 考虑 $\mathbb{R}^{n+1}$ 中的凸锥

$$K = \{(x, \lambda) \mid \lambda > 0, x \in \lambda M\}$$

以及线性变换 $B: \mathbb{R}^{n+1} \to \mathbb{R}^{m+1}$,

$$B(x, \lambda) = (Ax, \lambda).$$

由于 M 是闭的, 从引理 2.6.1 知

$$\mathrm{cl}\, K = 0^+(\mathrm{cl}\, K) = K \cup \{(z, 0) \mid z \in 0^+ M\}.$$

注意 B 的零空间 $\mathcal{N}(B) = \{(x, 0) \mid Ax = 0\}$, 故

$$0^+(\mathrm{cl}\, K) \cap \mathcal{N}(B) \subset (-0^+(\mathrm{cl}\, K)) \cap 0^+(\mathrm{cl}\, K).$$

于是根据上面已证明部分的结果有 $\mathrm{cl}(BK) = B(\mathrm{cl}\,K)$. 注意

$$\begin{aligned}&B(\mathrm{cl}\,K)\\ &= \{(Ax,\lambda) \mid \lambda > 0, x \in \lambda M\} \cup \{(Az,0) \mid z \in 0^+M\}.\end{aligned}$$

另一方面，由于 AM 是闭集，根据引理 2.6.1,

$$\begin{aligned}\mathrm{cl}(BK) &= \mathrm{cl}\{(y,\lambda) \mid \lambda > 0, y \in A(\lambda M) = \lambda AM\}\\ &= \{(y,\lambda) \mid \lambda > 0, y \in \lambda AM\} \cup \{(y,0) \mid y \in 0^+(AM)\}.\end{aligned}$$

于是 $\mathrm{cl}(BK) = B(\mathrm{cl}\,K)$ 意味着

$$\{Az \mid z \in 0^+M\} = 0^+(AM). \qquad \blacksquare$$

推论 2.6.3 设 $M_1,\cdots,M_m$ 为 $\mathbb{R}^n$ 中的非空凸集. 假定

$$z_i \in 0^+(\mathrm{cl}\,M_i),\ z_1+\cdots+z_m = 0 \Longrightarrow z_i \in L_i,\ 0 \le i \le m,$$

其中 L_i 为 $\mathrm{cl}\,M_i$ 的直线状子空间. 那么

$$\mathrm{cl}(M_1 + \cdots M_m) = \mathrm{cl}\,M_1 + \cdots + \mathrm{cl}\,M_m,$$

$$0^+(\mathrm{cl}(M_1 + \cdots M_m)) = 0^+(\mathrm{cl}\,M_1) + \cdots + 0^+(\mathrm{cl}\,M_m).$$

特别是如果 $M_1,\cdots M_m$ 均为闭集，则在上述假设下 $M_1 + \cdots M_m$ 是闭集.

证明 设 M 表示乘积欧氏空间 $\mathbb{R}^{mn} = \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ 中的集合 $M_1 \times \cdots \times M_m$, 并设线性变换 $A: \mathbb{R}^{mn} \to \mathbb{R}^n$,

$$A(x_1,\cdots,x_m) = x_1 + \cdots x_m, \quad \forall\, x_i \in \mathbb{R}^n.$$

那么 $AM = M_1 + \cdots + M_m$. 但显然,

$$\mathrm{cl}\,M = \mathrm{cl}\,M_1 \times \cdots \times \mathrm{cl}\,M_m,$$

因此不难看出,

$$0^+(\mathrm{cl}\, M) = 0^+(\mathrm{cl}\, M_1) \times \cdots \times 0^+(\mathrm{cl}\, M_m).$$

然后应用定理 2.6.2. ∎

推论 2.6.4 设 $K_1, \cdots . K_m$ 是 $\mathbb{R}^n$ 中的非空凸锥，满足 $z_i \in \mathrm{cl}\, K_i, z_1 + \cdots + z_m = 0 \Longrightarrow z_i \in L_i, i = 1, \cdots, m$, 这里 L_i 为 $\mathrm{cl}\, K_i$ 的直线状空间. 那么

$$\mathrm{cl}\,(K_1 + \cdots + K_2) = \mathrm{cl}\, K_1 + \cdots \mathrm{cl}\, K_m.$$

现在我们利用上述这些结果来讨论凸函数.

定理 2.6.5 设 $g : \mathbb{R}^n \to (-\infty, \infty]$ 为闭的真凸函数，并设 $A : \mathbb{R}^n \to \mathbb{R}^m$ 为线性变换. 假定

$$\{z \in \mathbb{R}^n \mid (g0^+)(z) \leq 0, (g0^+)(-z) > 0\} \cap \mathcal{N}(A) = \emptyset,$$

其中 $\mathcal{N}(A)$ 为 A 的零空间. 那么在 2.3 节中定义的函数 Ag, 即

$$(Ag)(y) = \inf\{g(x) \mid Ax = y\}, \quad \forall \in \mathbb{R}^m \tag{2.6.1}$$

是 $\mathbb{R}^m$ 上的一个闭的真凸函数，并且 $(Ag)0^+ = A(g0^+)$. 此外，当 $(Ag)(y) \neq \infty$ 时，式 (2.6.1) 右端中的下确界在某个 x 上达到.

证明 考虑非空闭凸集 $\mathrm{epi}\, g$ 和线性变换 $B : \mathbb{R}^{n+1} \to \mathbb{R}^{m+1}$, $B(x, \lambda) = (Ax, \lambda), x \in \mathbb{R}^n, \lambda \in \mathbb{R}$. 我们知道 $\mathrm{epi}\, g$ 的回收锥是 $\mathrm{epi}\,(g0^+)$, 并且 $\mathrm{epi}\, g$ 的直线状空间

$$\begin{aligned}&(-0^+(\mathrm{epi}\, g)) \cap 0^+(\mathrm{epi}\, g)\\ &= \{(z, \mu) \mid (g0^+)(z) \leq \mu, (g0^+)(-z) \leq -\mu\}.\end{aligned}$$

由此从假设可知 $\mathrm{epi}\, g$ 和 B 满足定理 2.6.2 的条件，得出结论 : $B(\mathrm{epi}\, g)$ 是一个非空闭凸集，其回收锥 $0^+(B(\mathrm{epi}\, g)) = B(\mathrm{epi}\,(g0^+))$. 此外，我们有

$$B(\operatorname{epi} g) = \operatorname{epi}(Ag),$$

$$B(\operatorname{epi}(g0^+)) = \operatorname{epi}(A(g0^+)).$$

于是由 $B(\operatorname{epi} g)$ 的闭性得知 Ag 的闭性，并且式 (2.6.1) 中的下确界当 $(Ag)(y) \neq \pm\infty$ 时在某个 x 处达到.

剩下要证 Ag 不取 $-\infty$. 为此只须证明其上图 $\operatorname{epi}(Ag)$ 不包含任何形如 $\{(z, \lambda) \mid \lambda \in I\!R\}$ 的铅垂直线. 但 $\operatorname{epi}(Ag)$ 中出现这样的直线意味着其回收锥

$$0^+(\operatorname{epi}(Ag)) = B(\operatorname{epi}(g0^+))$$

包含直线 $\{(0, \lambda) \mid \lambda \in I\!R\}$. 特别取 $\lambda < 0$ 可知 $(0, \lambda) \in \operatorname{epi}(A(g0^+))$, 从而必存在某个 $z \in I\!R^n$ 使得 $Az = 0$ 并且 $(z, \lambda) \in \operatorname{epi}(g0^+)$. 这样， $(g0^+)(z) \leq \lambda < 0$. 但 $g0^+$ 是正齐性凸函数，因此依据推论 2.2.9,

$$(g0^+)(-z) \geq -(g0^+)(z) > 0.$$

这与定理的假设相矛盾. ∎

推论 2.6.6 设 $f_1, \cdots, f_m$ 是 $I\!R^n$ 上的闭的真凸函数. 假定当 $z_1, \cdots, z_m \in I\!R^n$ 满足

$$\sum_{k=1}^{m}(f_k0^+)(z_k) \leq 0, \quad \sum_{k=1}^{m}(f_k0^+)(-z_k) > 0$$

时， $z_1 \cdots + z_m \neq 0$. 那么下端卷积 $f_1 \square \cdots \square f_m$ 也是 $I\!R^n$ 上的一个闭的真凸函数，并且 $(f_1 \square \cdots \square f_m)(x)$ 定义中的下确界对每个 x 都达到. 此外,

$$(f_1 \square \cdots \square f_m)0^+ = f_10^+ \square \cdots \square f_m0^+.$$

证明 如同推论 2.6.3 的证明那样，定义乘积空间 $\mathbb{R}^{mn} = \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ 到 $\mathbb{R}^n$ 的线性变换 A:

$$A(x_1, \cdots, x_m) = x_1 + \cdots + x_m, \quad x_i \in \mathbb{R}_n,$$

并设 g 为 $\mathbb{R}^{mn}$ 上的闭真凸函数:

$$g(x_1, \cdots, x_m) = f_1(x_1) + \cdots + f_m(x_m), \quad x_i \in \mathbb{R}^n.$$

然后应用定理 2.6.5 于 g 和 A 即得所要结论. ▮

定理 2.6.7 设 $f_1 \cdots f_m$ 为 $\mathbb{R}^n$ 上的闭真凸函数，假定 $f_1 + \cdots + f_m$ 不恒等于 ∞, 那么 $f_1 + \cdots + f_m$ 也是闭的真凸函数，并且

$$(f_1 + \cdots + f_m)0^+ = f_1 0^+ + \cdots + f_m 0^+.$$

此外，若诸 f_i 不一定都是闭的，但 $\bigcap_{k=1}^{m} \mathrm{ri}\,(\mathrm{dom}\, f_i) \neq \emptyset$, 则

$$\mathrm{cl}\,(f_1 + \cdots + f_m) = \mathrm{cl}\, f_1 + \cdots + \mathrm{cl}\, f_m.$$

证明 令 $f = f_1 + \cdots + f_m$, 并设

$$x \in \mathrm{ri}\,(\mathrm{dom}\, f) = \mathrm{ri}\,(\bigcap_{i=1}^{m} \mathrm{dom}\, f_i).$$

依据定理 2.4.11, 对于每个 $y \in \mathbb{R}^n$,

$$(\mathrm{cl}\, f)(y) \lim_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y) = \sum_{i=1}^{m} \lim_{\lambda \uparrow 1} f_i((1-\lambda)x + \lambda y).$$

如果每一 f_i 都是闭的，则 $\lim_{\lambda \uparrow 1} f_i((1-\lambda)x + \lambda y) = f_i(y)$. 从而 $(\mathrm{cl}\, f)(y) = f(y)$; 而如果

$$\bigcap_{i=1}^{m} \mathrm{ri}\,(\mathrm{dom}\, f_i) \neq \emptyset,$$

则依定理 1.4.13,

$$\mathrm{ri}\,(\mathrm{dom}\, f) = \bigcap_{i=1}^{m} \mathrm{ri}\,(\mathrm{dom}\, f_i),$$

可见 $x \in \mathrm{ri}\,(\mathrm{dom}\, f_i), i = 1, \cdots, m$. 然后再利用定理 2.4.11 得

$$(\mathrm{cl}\, f_i)(y) = \lim_{\lambda \uparrow 1} f_i((1-\lambda)x + \lambda y),$$

从而 $\mathrm{cl}\, f = \mathrm{cl}\, f_1 + \cdots + \mathrm{cl}\, f_m$. 有关 $f0^+$ 的公式从定理 2.3.10 推出. ∎

定理 2.6.8 设 $g: \mathbb{R}^m \to (-\infty, \infty]$ 为真凸函数, 并设 $A: \mathbb{R}^n \to \mathbb{R}^m$ 为线性变换. 假定 gA 不恒等于 ∞. 如果 g 是闭的, 则 gA 也是闭的, 并且 $(gA)0^+ = (g0^+)A$. 而如果 g 不闭, 但存在 $x \in \mathbb{R}^n$ 使得 $Ax \in \mathrm{ri}\,(\mathrm{dom}\, g)$, 则 $\mathrm{cl}\,(gA) = (\mathrm{cl}\, g)A$.

证明 我们知道 gA 是真凸函数. 如果定义线性变换 $B: \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$,

$$B(x, \lambda) = (Ax, \lambda), \quad \forall\, x \in \mathbb{R}^n, \lambda \in \mathbb{R},$$

则 $\mathrm{epi}\,(gA) = B^{-1}\mathrm{epi}\, g$. 因此当 g 闭时 gA 也是闭的, 并且从推论 2.3.8 直接可得 $(gA)0^+ = (g0^+)A$. 现在若 g 不闭, 但 $Ax \in \mathrm{ri}\,(\mathrm{dom}\, g)$, 则从引理 2.4.2 知, 只要取 $\mu > g(Ax)$ 就有 $B(x, \mu) = (Ax, \mu) \in \mathrm{ri}\,(\mathrm{epi}\, g)$, 从而 $B^{-1}(\mathrm{ri}\,(\mathrm{epi}\, g)) \neq \emptyset$. 于是应用定理 1.4.17, 得到

$$B^{-1}(\mathrm{cl}\,(\mathrm{epi}\, g)) = \mathrm{cl}\,(B^{-1}(\mathrm{epi}\, g)),$$

这正是说 $\mathrm{cl}\,(gA) = (\mathrm{cl}\, g)A$. ∎

定理 2.6.9 设 $\{f_i \mid i \in I\}$ 是 $\mathbb{R}^n$ 上一族真凸函数, I 为任意指标集. 设

$$f = \sup\{f_i \mid i \in I\}.$$

假定 f 在某点取有穷值，并且每一 f_i 是闭的，那么 f 也是闭真凸的，并且

$$f0^+ = \sup\{f_i0^+ \mid i \in I\}.$$

如果 f_i 不一定全是闭的，但存在一点 $z \in \bigcap_{i\in I} \mathrm{ri}\,(\mathrm{dom}\, f_i)$ 使得 $f(z)$ 有穷，那么

$$\mathrm{cl}\, f = \sup\{\mathrm{cl}\, f_i \mid i \in I\}.$$

证明 由于 $\mathrm{epi}\, f$ 是诸集合 $\mathrm{epi}\, f_i$ 之交，故当每个 f_i 为闭的时候 f 也是闭的. 关于 $f0^+$ 的公式是推论 2.3.8 之 (3) 的结果. 注意诸集合 $\mathrm{ri}\,(\mathrm{epi}\, f_i)$ 的交包含点 $(z, f(z)+1)$ (见引理 2.4.2), 于是 $\mathrm{cl}\, f$ 的公式从定理 1.4.13 得到. ▮

定理 2.6.10 设 $\mathbb{R}^n$ 中非空闭凸集 $M_1, \cdots, M_m$ 满足如下条件: $z_i \in 0^+M_i, z_1 + \cdots + z_m = 0 \Longrightarrow z_i$ 属于 M_i 的直线状空间, $i = 1, \cdots, m$. 记 $M = \mathrm{conv}\,(M_1 \cap \cdots \cap M_m)$ 那么,

$$\mathrm{cl}\, M = \cup\left\{\sum_{i=1}^{m} \lambda_i M_i \mid \lambda_i \geq 0^+, \sum_{i=1}^{m} \lambda_i = 1\right\}.$$

这里记号 $\lambda_i \geq 0^+$ 意指当 $\lambda_i = 0$ 时 $\lambda_i M_i$ 取作 0^+M_i 而不是 $\{0\}$. 此外,

$$0^+(\mathrm{cl}\, M) = 0^+M_1 + \cdots + 0^+M_m.$$

证明 设 K_i 是 $\mathbb{R}^{n+1}$ 中由 $\{(x_i, 1) \mid x_i \in M_i\}$ 生成的凸锥, $i = 1, \cdots, m$. 从引理 2.6.1 知,

$$\mathrm{cl}\, K_i = K_i \cup \{(x, 1) \mid x \in 0^+M_i\}$$

于是根据关于回收锥 0^+M_i 的假设, 应用推论 2.6.4 得到

$$\mathrm{cl}\,(K_1 + \cdots + K_m) = \mathrm{cl}\, K_1 + \cdots + \mathrm{cl}\, K_m.$$

注意 $\text{cl}\,(K_1+\cdots K_m)$ 与 $I\!R^{n+1}$ 中超平面 $H_1=\{(x,1)\mid x\in I\!R^n\}$ 之交是 $K_1+\cdots+K_m$ 与 H_1 之交的闭包，而 $K_1+\cdots+K_m$ 与 H_1 之交恰好由形如 $(x,1)$ 之组成，其中 x 属于某个凸组合 $\lambda_1 M_1+\cdots+\lambda_m M_m$. 但所有这样的凸组合之并正好是 M (见习题 1.3.18), 因此,

$$\begin{aligned}\text{cl}\,(K_1+\cdots+K_m)\cap H_1&=\{(x,1)\mid x\in\text{cl}\,M\}\\&=(\text{cl}\,M_1+\cdots+\text{cl}\,M_m)\cap H_1.\end{aligned}$$

这正是说

$$\text{cl}\,M=\cup\Big\{\sum_{i=1}^m\lambda_i M_i\mid\lambda_i\geq 0^+,\ \sum_{i=1}^m\lambda_i=1\Big\}.$$

由此可见 $\text{cl}\,(K_1+\cdots+K_m)$ 是 $I\!R^{n+1}$ 中由 $\{(1,x)\mid x\in\text{cl}\,M\}$ 生成的凸锥的闭包，从而它所包含的形如 $(x,0)$ 的向量中的 $x\in 0^+(\text{cl}\,M)$(见引理 2.6.1). 另一方面，包含在 $\text{cl}\,K_1+\cdots+\text{cl}\,K_m$ 中的形如 $(x,0)$ 的向量中的 $x\in 0^+M_1+\cdots+0^+M_m$. 这样证明了关于 $0^+(\text{cl}\,M)$ 的公式成立. ▮

推论 2.6.11 设 $M_1,\cdots,M_m$ 是 $I\!R^n$ 中的非空闭凸集，假定它们有相同的回收锥 K. 那么凸集 $M=\text{conv}\,(M_1\cup\cdots\cup M_m)$ 是闭的并且其回收锥也是 K.

证明 设 $z_i\in K$ 满足 $z_1+\cdots+z_m=0$, 则

$$-z_1=z_2+\cdots+z_m\in(-K)\cap K.$$

类似地可知 $z_i\in(-K)\cap K, i=1,\cdots,m$, 即 z_i 属于 M_i 的直线状空间，从而定理 2.6.10 适用于目前的情形. 注意到 $0^+M_i=K, i=1,\cdots,m$, 可知当 $\lambda_j>0$ 时

$$0^+M_i+\lambda_j M_j=\lambda_j(K+M_j)=\lambda_j M_j=0M_i+\lambda_j M_j,$$

这表明定理 2.6.10 中的 $\lambda_i \geq 0^+$ 可以写成 $\lambda_i \geq 0$. 因此 $\mathrm{cl}\, M = M$. ∎

推论 2.6.12 设 $M_1, \cdots, M_m$ 是 $\mathbb{R}^n$ 中的有界闭凸集，则 $\mathrm{conv}\,(M_1 \cup \cdots \cup M_m)$ 也是闭有界的.

证明 这是因为对于非空 $M_i, 0^+ M_i = \{0\}$. ∎

推论 2.6.13 设 $f_1 \cdots, f_m$ 是 $\mathbb{R}^n$ 上闭真凸函数，假定它们都有相同的回收函数 k. 那么 $f = \mathrm{conv}\,\{f_1, \cdots, f_m\}$ 也是闭真凸的，并且其回收函数也是 k. 此外，在定理 2.3.5 中关于 $f(x)$ 的表达式 (2.3.1) 中的下确界在某个凸组合向量处达到.

证明 在推论 2.6.11 中取 $M_i = \mathrm{epi}\, f_i$, 根据假设，$0^+ M_i = \mathrm{epi}\,(f_i 0^+) = \mathrm{epi}\, k = K$. 因此 M 为 $\mathbb{R}^{n+1}$ 中的闭凸集，并且 $0^+ M = K$. 显然 M 必定是某个闭真凸函数 g 的上图，即 $M = \mathrm{epi}\, g$. 从定理 2.3.5 不难看出 $f = g$. ∎

习 题 2.6

2.6.1 设 f_1, f_2 是 $\mathbb{R}^n$ 上的闭真凸函数. 假定

$$(f_1 0^+)(z) + (f_2 0^+)(-z) > 0, \quad \forall\, z \neq 0.$$

试证: $f_1 \Box f_2$ 是闭真凸函数，并且在下确界公式

$$(f_1 \Box f_2)(x) = \inf\{f_1(x - y) + f_2(y) \mid y \in \mathbb{R}^n\}$$

中的下确界对任意固定的 x 都被某个 y 达到.

2.6.2 设 $f = f_2$ 为 $\mathbb{R}^n$ 的闭真凸函数，$f_1 = \delta(\cdot \mid -M)$ 为 $-M$ 的示性函数，这里 M 为 $\mathbb{R}^n$ 的非空闭凸子集. 试证:

$$(f_1 \Box f_2)(x) = \inf\{f(y) \mid y \in (M + x)\};$$

此外，如果 f 和 M 没有公共回收方向，则 $f_1 \Box f_2$ 是下半连续的，并且对于每一 $x \in \mathbb{R}^n$，上述下确界被达到.

2.6.3 设 M 是 $\mathbb{R}^n$ 的非负象限，并且对于 $x, y \in \mathbb{R}^n$，$y \geq x$ 是指 $y - x \in M$. 设 f 是 $\mathbb{R}^n$ 上的闭真凸函数，其回收锥不包含任何非负非零向量. 试证：

$$g(x) = \inf\{f(y) \mid y \geq x\}, \quad \forall x \in \mathbb{R}^n$$

是 $\mathbb{R}^n$ 上的闭真凸函数，并且上述下确界对一切 $x \in \mathbb{R}^n$ 都被某个 $y \geq x$ 所达到.

第三章　对偶关系

数学的许多分枝中对偶性是一个十分重要的概念. 例如, 复数的共轭, 矩阵的转置或共轭转置, 微分方程和伴随方程, 控制理论中的能控性与能观测性, 等等. 凸分析中也同样充满着种种对偶关系. 本章中我们将介绍凸函数的共轭, 凸集的承托函数, 凸集的极化及对偶锥, 等等. 通过这些概念的引入和深入研究, 将不断深化凸分析中研究对象的认识, 从而揭示其内在的一些本质联系.

3.1　凸函数的共轭函数

在上一章中我们曾指出, 每个闭真凸函数 $f: \mathbb{R}^n \to (-\infty, \infty]$ 都可以表示成仿射函数族 $A(f)$ 的上包络:

$$f(x) = \sup\{a(x) \mid a \in A(f)\}, \quad \forall x \in \mathbb{R}^n,$$

这里 $A(f)$ 表示不超过 f 的仿射函数全体. $A(f)$ 中的每个仿射函数 a 都可以表示成

$$a(x) = \langle x^*, x\rangle - \mu,$$

其中 $x^* \in \mathbb{R}^n$, $\mu \in \mathbb{R}$. 容易看出,

$$a \in A(f) \Longleftrightarrow \mu \geq \sup\{\langle x^*, x\rangle - f(x) \mid x \in \mathbb{R}^n\}.$$

对于 $\mathbb{R}^n$ 上的凸函数 f, 我们定义一个新的函数 f^*:

$$f^*(x^*) = \sup\{\langle x^*, x\rangle - f(x) \mid x \in \mathbb{R}^n\}, \quad \forall x^* \in \mathbb{R}^n. \tag{3.1.1}$$

这个新的函数 f^* 叫做 f 的共轭函数. f^* 的共轭函数

$$f^{**}(x) = \sup\{\langle x, x^*. - f^*(x) \mid x^* \in I\!R^n\}, \quad \forall \in I\!R^n \tag{3.1.2}$$

则称为 f 的二次共轭函数.

把 f 变成 f^* 的变换称为 Lagrange-Fenchel 变换. 这种变换起源于力学. 在理论力学中通常通过这样的变换把 Lagrange 函数变成 Hamilton 函数. f 与 f^* 的这种对偶关系在经济学中也有直接的解释. 例如，设 x 表示产出丛，x^* 表示产出价格系，f 为成本函数，则

$$f^*(x^*) = \sup\{\langle x^*, x\rangle - f(x) \mid x \in I\!R^n\}$$

就是利润函数，即在产出价格系为 x^* 时企业所能获得的最大利润. 有兴趣的读者可参阅数理经济学的专著.

当 f 是 $I\!R^n$ 上的真凸函数时，$f(x) > -\infty, \forall x \in I\!R^n$, 并且存在 $x_0 \in I\!R^n$ 使得 $f(x_0) < \infty$. 于是从定义 (3.1.1) 可以看出 $f^*(x^*) > -\infty, \forall x^* \in I\!R^n$. 又由定理 2.4.13 的证明知 $A(f) \neq \emptyset$, 即存在 $x_0* \in I\!R^n$ 和 $\mu \in I\!R$ 使

$$f(x) \geq \langle x_0^*, x\rangle - \mu, \quad \forall x \in I\!R^n,$$

即

$$\langle x_0^*, x\rangle - f(x) \leq \mu, \quad \forall x \in I\!R^n.$$

于是

$$f^*(x_0^*) \leq \mu < \infty.$$

因此 f^* 也是 $I\!R^n$ 上的真凸函数 (f^* 作为一族仿射函数的上包络当然是凸的). 此外，从定义 (3.1.1) 看出 f^* 还是闭的，不管 f 本身闭与否. 这就是说，当 f 是 $I\!R^n$ 上的真凸函数时，其共轭 f^* 总是 $I\!R^n$ 上的闭真凸函数. 同样，f 的二次共轭 f^{**} 也一定是 $I\!R^n$ 上的闭的真凸函数.

另一方面，当存在一点 $x_0 \in \mathbb{R}^n$ 使得 $f(x_0) = -\infty$ 时，$f^*(x^*) \equiv \infty, \forall x^* \in \mathbb{R}^n$; 从而 $f^{**}(x) \equiv -\infty, \forall x \in \mathbb{R}^n$.

从定义直接看出，对于 $\mathbb{R}^n$ 上的两个凸函数 f, g, 当 $f \leq g$ 时，有 $g^* \leq f^*$.

定理 3.1.1 (Young不等式) 设 $f : \mathbb{R}^n \to (-\infty, \infty]$ 为真凸函数，那么

$$f(x) + f^*(x^*) \geq \langle x^*, x \rangle, \quad \forall x \in \mathbb{R}^n, \forall x^* \in \mathbb{R}^n. \quad (3.1.3)$$

证明 从 f^* 的定义 (3.1.1) 直接得到. ▮

定理 3.1.2 设 f 为 $\mathbb{R}^n$ 上的凸函数，那么

$$f^* = (\mathrm{cl}\, f)^*, \ f^{**} = \mathrm{cl}\, f.$$

特别 $f^{**} = f$, 当且仅当 f 为闭的真凸函数，或 f 为常值函数 $+\infty$ 和 $-\infty$.

证明 首先注意，常值函数 $+\infty$ 和 $-\infty$ 互为共轭，并且这两个函数是仅有的非真的闭凸函数. 因此我们只须考虑真凸函数 f 的情况.

如果 $f^{**} = f$, 则 f 是闭的. 反之，如果 f 是闭的，则依据定理 2.4.15, f 是仿射函数族 $A(f)$ 的上包络，而且由本节开头的分析,

$$a \in A(f), \quad a(x) = \langle x^*, x \rangle - \mu \iff \mu \geq f^*(x^*).$$

于是对于任意 $x \in \mathbb{R}^n$,

$$\begin{aligned} f(x) &= \sup\{a(x) \mid a \in A(f)\} \\ &= \sup\{\langle x^*, x \rangle - \mu \mid x^* \in \mathbb{R}^n, \mu \geq f^*(x^*)\} \\ &= \sup\{\langle x^*, x \rangle - f^*(x^*) \mid x^* \in \mathbb{R}^n\} \\ &= f^{**}(x), \end{aligned}$$

即 $f^{**}=f$.

另一方面，若 f 是真凸函数，则 $A(f)=A(\mathrm{cl}\, f)$. 事实上，对于 $I\!R^n$ 上的仿射函数 $a(x)$,

$$a(x)\le f(x),\quad \forall\, x\in I\!R^n \Longleftrightarrow a(x)\le \mathrm{cl}\, f(x),\quad \forall\, x\in I\!R^n.$$

从而由定义 (3.1.1) 立即推出

$$f^*(x^*)=(\mathrm{cl}\, f)^*(x^*),\quad \forall\, x^*\in I\!R^n.$$ ▮

推论 3.1.3 共轭运算 $f\to f^*$ 在 $I\!R^n$ 上的一切闭真凸函数类中是一种一一对应关系.

下面讨论共轭函数的一些例子.

首先考虑 $I\!R$ 上的闭真凸函数 $f(x)=e^x$. 由定义,

$$f^*(x^*)=\sup\{xx^*-e^x \mid x\in I\!R\},\quad \forall\, x^*\in I\!R.$$

如果 $x^*<0$, 则 x 取负值可以使 xx^*-e^x 任意大. 从而其上确界为 $+\infty$; 如果 $x^*>0$, 通过微分判别法容易验证，当 $x=\ln x^*$ 时， xx^*-e^x 达到最大值 $x^*\ln x^*-x^*$；而当 $x^*=0$ 时显然上确界为 0 . 于是指数函数 e^x 的共轭函数为

$$f^*(x^*)=\begin{cases} x^*\ln x^*-x^*, & x^*>0,\\ 0, & x^*=0,\\ +\infty, & x^*<0.\end{cases}$$

f 的二次共轭函数 f^{**} 也可直接计算:

$$\begin{aligned} f^{**} &= \sup\{xx^*-f^*(x^*)\mid x^*\in I\!R\}\\ &= \sup\{xx^*-x^*\ln x^*+x^*\mid x^*>0\}\\ &= e^x,\quad \forall\, x\in I\!R.\end{aligned}$$

从这个例子我们看到，一个处处取有穷值的凸函数未必有处处取有穷值的共轭函数.

我们再来讨论 $I\!R^n$ 中的凸函数 f:

$$f(x)=\begin{cases}\frac{1}{2}\|x\|^2, & \|x\|\le a,\\ +\infty, & \|x\|>a,\end{cases}$$

其中 $a>0$ 为给定的数. 根据定义，对于 $x^*\in I\!R^n$,

$$f^*(x^*)=\sup\left\{\langle x^*,x\rangle-\frac{1}{2}\|x\|^2 \;\Big|\; \|x\|\le a\right\}.$$

但是

$$\langle x^*,x\rangle-\frac{1}{2}\|x\|^2=\frac{1}{2}\|x^*\|^2-\frac{1}{2}\|x-x^*\|^2,$$

而且不难看出，对于任意 $x^*\in I\!R^n$, $\|x^*\|\ge a$, 仅当 $x=x_0\triangleq ax^*/\|x^*\|$ 时，$\|x-x^*\|^2$ 在闭球 $\overline{B}(0,a)$ 上达到最小值，即

$$\|x_0-x^*\|^2=\inf\{\|x-x^*\|^2 \mid x\in\overline{B}(0,a)\}.$$

因此，这时函数 $\langle x^*,x\rangle-\frac{1}{2}\|x\|^2$ 当 $x=ax^*/\|x^*\|$ 时，闭球 $\overline{B}(0,a)$ 上达到最大值 $a\|x^*\|-\frac{1}{2}a^2$. 而当 $\|x^*\|<a$ 时，显然 $f^*(x^*)=\frac{1}{2}\|x^*\|^2$. 因此,

$$f^*(x^*)=\begin{cases}\frac{1}{2}\|x^*\|^2, & \forall x^*\in I\!R^n, \|x^*\|<a,\\ a\|x^*\|-\frac{1}{2}a^2, & \forall x^*\in I\!R^n, \|x^*\|\ge a.\end{cases}$$

下面我们列出一些闭真凸函数的共轭双，这里采用记号： $p>1$, $\frac{1}{p}+\frac{1}{q}=1$, $\|x\|_p$ 表示 $I\!R^n$ 上的 p 范数,

$$\|x\|_p=\left(\sum_{i=1}^n|\xi_i|^p\right)^{\frac{1}{p}},\quad \forall\, x=(\xi_1,\cdots,\xi_n)^2\in I\!R^n.$$

例 3.1.1 $f(x)=\|x\|_p^p/p$, $\forall\, x\in I\!R^n$, $1<p<\infty$,

$$f^*(x^*) = \|x^*\|_q^q, \quad \forall\, x^* \in I\!R^n,\ \frac{1}{p} + \frac{1}{q} = 1.$$

例 3.1.2 $f(x) = \langle x_0^*, x\rangle - \alpha, \quad \forall\, x \in I\!R^n,$

$$f^*(x^*) = \begin{cases} \alpha, & x^* = x_0^*, \\ +\infty, & \forall x^* \in I\!R^n \backslash \{x_0^*\}. \end{cases}$$

例 3.1.3 对于 $0 < p < 1,\ 1/p + 1/q = 1$,

$$f(x) = \begin{cases} -x^p/p, & \forall x \geq 0, \\ +\infty, & \forall x < 0, \end{cases}$$

$$f^*(x^*) = \begin{cases} -|x^*|_q^q/q, & \forall x^* < 0, \\ +\infty, & \forall x^* \geq 0. \end{cases}$$

例 3.1.4

$$f(x) = \begin{cases} -\frac{1}{2} - \ln x, & \forall x > 0, \\ +\infty, & \forall x^* \leq 0, \end{cases}$$

$$f^*(x^*) = \begin{cases} -\frac{1}{2} - \ln(-x^*), & \forall x^* < 0, \\ +\infty, & \forall x^* \geq 0, \end{cases}$$

例 3.1.5 对于 $a > 0$,

$$f(x) = \begin{cases} -(a^2 - \|x\|^2)^{1/2}, & \forall \|x\| \leq a, \\ +\infty, & \forall \|x\| > a, \end{cases}$$

$$f^*(x^*) = a(1 + \|x^*\|^2)^{1/2}, \quad \forall\, x^* \in I\!R^n.$$

例 3.1.6 $f(x) = \|x\|, \forall x \in I\!R^n$. 我们较详细地来计算 f 的共轭. 当 $\|x^*\| > 1$ 时, 取 $x = \alpha x^*, \alpha > 0$. 于是

$$\langle x, x^*\rangle - \|x\| = \alpha(\|x^*\|^2 - \|x^*\|).$$

由此可知

$$f^*(x^*) = \sup\{\langle x, x^*\rangle - \|x\| \mid x \in \mathbb{R}^n\} = +\infty.$$

另一方面，当 $\|x^*\| \le 1$ 时

$$\langle x, x^*\rangle - \|x\| \le \|x\|(\|x^*\| - 1) \le 0.$$

特别当 $x = 0$ 时，

$$\langle x, x^*\rangle - \|x\| = 0.$$

因此 $f^*(x^*) = 0$. 这样我们得到

$$f^*(x^*) = \begin{cases} 0, & \|x^*\| \le 1, \\ +\infty, & \|x^*\| > 1. \end{cases}$$

或者

$$f^*(x^*) = \delta(x^*|\overline{B}(0,1)),$$

其中 $\overline{B}(0,1)$ 是 $\mathbb{R}^n$ 中的闭单位球.

在例 3.1.4 中我们有 $f^*(x^*) = f(-x^*)$. 实际上有不少凸函数都有这个性质. 但在 $\mathbb{R}^n$ 上方程 $f^* = f$ 却只有唯一解. 事实上从例 3.1.1 中我们就看到，如果令 $u(x) = \frac{1}{2}\|x\|^2$, $\forall x \in \mathbb{R}^n$, 则 $u^* = u$. 今若 $\mathbb{R}^n$ 上的一个凸函数 f 满足 $f^* = f$, 则 f 当然是闭的真凸函数. 又依 Young 不等式，

$$\langle x, x\rangle \le f(x) + f^*(x) = 2f(x).$$

因此 $u \le f$. 从这个不等式我们又可得到 $f^* \le u^*$. 但 $f^* = f$, $u^* = u$, 所以 $f \le u$, 即 $f = u$.

设 L 是 $\mathbb{R}^n$ 的子空间， f 表示 L 的示性函数，即 $f(x) = \delta(x|L)$, $x \in \mathbb{R}^n$. 那么 f 的共轭函数是

$$\begin{aligned} f^*(x^*) &= \sup\{\langle x^*, x\rangle - \delta(x|L) \mid x \in \mathbb{R}^n\} \\ &= \sup\{\langle x^*, x\rangle - \delta(x|L) \mid x \in L\}. \end{aligned}$$

如果 $x^* \in L^\perp$, 即 $\langle x^*, x\rangle = 0$, $\forall x \in L$, 则 $f^*(x^*) = 0$; 而

如果 $x^* \notin L^{\perp}$, 则显然 $f^*(x^*) = \infty$. 因此 f^* 是子空间 L 的正交补 $L^{\perp}$ 的示性函数. 关系式 $f^{**} = f$ 在这里对应于 $L^{\perp\perp} = L$.

上述这种子空间的正交对应关系也可以适当地推广到仿射集. 我们知道 $\mathbb{R}^n$ 的任一仿射集 M 都可以表示成 $M = L+a$, 其中 L 为 $\mathbb{R}^n$ 的子空间, $a \in \mathbb{R}^n$. 我们称 f 是 M 上的部分仿射函数, 是指它是一个真凸函数, 并且 $\operatorname{dom} f = M$, f 在 M 上是仿射的. 显然 f 可以表示成

$$f(x) = \delta(x|L+a) + \langle x, a^* \rangle + \alpha, \tag{3.1.4}$$

其中 $a^* \in \mathbb{R}^n, \alpha \in \mathbb{R}$. 不难算出, f 的共轭函数 f^* 仍然是一个部分仿射函数:

$$f^*(x^*) = \delta(x^*|L^{\perp} + a^*) + \langle x^*, a \rangle + \alpha^*, \tag{3.1.5}$$

其中 $\alpha^* = -\alpha - \langle a, a^* \rangle$. 我们讨论更一般的情形.

定理 3.1.4 设 $g : \mathbb{R}^n \to (-\infty, \infty]$ 为真凸函数, 并令

$$f(x) = g(A(x-a)) + \langle x, a^* \rangle + \alpha, \quad \forall\, x^* \in \mathbb{R}^n,$$

其中 A 是 $\mathbb{R}^n$ 到 $\mathbb{R}^n$ 上的一对一的线性变换, a 和 a^* 为 $\mathbb{R}^n$ 的向量, $\alpha \in \mathbb{R}$. 那么对于 $x^* \in \mathbb{R}^n$,

$$f^*(x^*) = g^*(A^{*-1}(x^* - a^*)) + \langle x^*, a \rangle + \alpha^*, \tag{3.1.6}$$

其中 A^* 为 A 的伴随, 而 $\alpha^* = -\alpha - \langle a, a^* \rangle$.

证明 作变量替换 $y = A(x-a)$, 从共轭函数的定义我们得到

$$\begin{aligned} & f^*(x^*) \\ = & \sup\{\langle x^*, x \rangle - g(A(x-a)) - \langle x, a^* \rangle - \alpha \mid x \in \mathbb{R}^n\} \\ = & \sup\{\langle A^{-1}y + a, x^* - a^* \rangle - g(y) - \alpha \mid y \in \mathbb{R}^n\} \end{aligned}$$

$$= \sup\{\langle y, A^{*-1}(x^* - a^*)\rangle - g(y) \mid y \in \mathbb{R}^n\} + \langle x^*, a\rangle + \alpha^*$$
$$= g^*(A^{*-1}(x^* - a^*)) + \langle x^*, a\rangle + \alpha^*. \quad \blacksquare$$

特别在上述定理中取 $g(x) = \delta(x|L)$, $A = I$, 式 (3.1.6) 就变换成式 (3.1.5).

利用定理 3.1.4 可以得到 $\mathbb{R}^n$ 上一般的二次型凸函数的共轭函数. 这里关键是要找出二次型函数

$$g(x) = \frac{1}{2}\langle Qx, x\rangle, \quad \forall\, x \in \mathbb{R}^n$$

的共轭函数, 其中 Q 为 $n \times n$ 非负定矩阵. 如果 Q 是非奇异的, 则容易算出 $\langle x, x^*\rangle - g(x)$ 的上确界在 $x = A^{-1}x^*$ 处达到, 故

$$g^*(x^*) = \frac{1}{2}\langle A^{-1}x^*, x^*\rangle, \quad \forall\, x^* \in \mathbb{R}^n.$$

如果 Q 是奇异矩阵, 并记 $L = \{Qx \mid x \in \mathbb{R}^n\}$, 则 $L^\perp = \{x \in \mathbb{R}^n \mid Qx = 0\}$, 而且 $\mathbb{R}^n$ 有正交分解 $\mathbb{R}^n = L + L^\perp$. 设 Q 在 L 上的限制记作 Q_L, 则 Q_L 在 L 上是对称非奇异的. 我们把 $x \in \mathbb{R}^n$ 分解成

$$x = y + z, \quad y \in L,\ z \in L^\perp.$$

于是

$$g^*(x^*) = \sup\left\{\langle x^*, y\rangle + \langle x^*, z\rangle - \frac{1}{2}\langle Q_L y, y\rangle \,\middle|\, y \in L, z \in L^\perp\right\}.$$

由此容易算出

$$g^*(x^*) = \begin{cases} \frac{1}{2}\langle Q_L^{-1}x^*, x^*\rangle, & \forall x^* \in L, \\ +\infty, & \forall x^* \notin L. \end{cases}$$

设 $f: \mathbb{R}^n \to (-\infty, \infty]$ 是一个闭的真凸函数, 于是 $f^{**} = f$. 根据定义,

$$f^*(0) = \sup\{\langle 0, x\rangle - f(x) \mid x \in I\!R^n\}$$
$$= -\inf\{f(x) \mid x \in I\!R^n\},$$

$$f(0) = f^{**}(0) = \sup\{-f^*(x^*) \mid x^* \in I\!R^n\}$$
$$= -\inf\{f^*(x^*) \mid x^* \in I\!R^n\}.$$

由此可见，关系式

$$\inf\{f(x) \mid x \in I\!R^n\} = 0 = f(0)$$

成立的充分必要条件是

$$\inf\{f^*(x^*) \mid x^* \in I\!R^n\} = 0 = f^*(0).$$

换句话说，如果把在原点达到最小值 0 的非负闭真凸函数归入一类，则这一类函数在 Lagrange-Fenchel 变换之下是不变的.

称 $I\!R^n$ 上的一个闭凸函数 f 为对称的，是指它满足

$$f(-x) = f(x), \quad \forall\, x \in I\!R^n.$$

显然，一个闭凸函数 f 为对称的充要条件是其共轭函数 f^* 也必是对称的. 其实，这一对称性是更一般的对称性结果的特殊情形. 设 G 表示 $I\!R^n$ 上一些正交线性变换组成的集合，一个函数 $f : I\!R^n \to (-\infty, \infty]$ 称为相对于 G 是对称的，是指它满足

$$f(Ax) = f(x), \quad \forall\, x \in I\!R^n, \forall A \in G.$$

于是通常的对称性对应于 G 由单个变换 $A : x \to -x$ 组成的情形.

定理 3.1.5 设 f 为 $\mathbb{R}^n$ 上的闭真凸函数，G 为 $\mathbb{R}^n$ 中一些正交线性变换组成的集合. 那么 f 相对于 G 是对称的充分必要条件是：f^* 相对 G 也是对称的.

证明 在定理 3.1.4 中取 $g=f$, $a=a^*=0$, $\alpha=0$, 从 $fA=f$ 我们推出 $f^*(A^{*-1})=f^*$. 但是对于正交变换 $A\in G$, $A^{*-1}=A$. 因此 $f^*A=f^*, \forall A\in G$. 但 f 是闭真凸函数，故 $f^{**}=f$. 于是从 $f^*A=f^*, \forall A\in G$ 同样推出 $fA=f, \forall A\in G$. ∎

习 题 3.1

3.1.1 验证本节中列举的几个闭真凸函数共轭双的例子.

3.1.2 我们称 $\mathbb{R}^n$ 上的真凸函数 f 为部分二次型凸函数，是指 f 可以表示成

$$f(x)=g(x)+\delta(x|M),\ x\in\mathbb{R}^n,$$

其中 g 为 $\mathbb{R}^n$ 上的有穷值二次型凸函数，M 为 $\mathbb{R}^n$ 的一个仿射集, 例如,

$$f(x)=\frac{1}{2}(\lambda_1\xi_1^2+\cdots+\lambda_n\xi_n^2),\quad \forall\, 0\le\lambda_j\le+\infty$$

是一个部分二次型凸函数,

$$\operatorname{dom} f=\{x=(\xi_1,\cdots,\xi_n)^2 \mid \xi_i=0, \forall j \text{ 使得 } \lambda_j=+\infty\}.$$

试证：

$$f^*(x^*)=\frac{1}{2}(\lambda_1\xi_1^{*2}+\cdots+\lambda_n^*\xi_n^{*2}),\ 0\le\lambda_j^*\le+\infty,$$

其中 $\lambda_j^*=1/\lambda_j$ (这是 $1/\infty$ 理解成 0, 而 $1/0$ 理解成 ∞).

3.1.3 给定 $a_1,a_2\in\mathbb{R}^n$ 和 $\beta_1,\beta_2\in\mathbb{R}$, 设

$$f(x)=\max\{\langle x,a_1\rangle+\beta_1,\langle x,a_2\rangle+\beta_2\},\ x\in\mathbb{R}^n.$$

试证：

$$f^*(x^*)=\begin{cases}-\lambda_1\beta_1-\lambda_2\beta_2, & x^*=\lambda_1a_1+\lambda_2a_2,\\ +\infty, & x^*\notin \operatorname{span}\{a_1,a_2\}\end{cases}.$$

3.1.4 给定 $a\in\mathbb{R}^n$ 和 $\alpha,\beta\in\mathbb{R}$, 设

$$f(x)=\begin{cases}\beta, & \text{如果 } \langle x,a\rangle\le\alpha,\\ +\infty, & \text{如果 } \langle x,a\rangle>\alpha.\end{cases}$$

试证:

$$f^*(x^*) = \begin{cases} \lambda\alpha - \beta, & \text{如果 } x^* = \lambda a, \text{ 且 } \lambda \geq 0, \\ +\infty, & \text{其余处}. \end{cases}$$

3.1.5 设 $g : I\!R^n \times I\!R^m \to (-\infty, +\infty]$ 为真凸函数, 定义 $f(x) = \inf\{g(x,y) \mid y \in I\!R^n\}$. 试证: $f^*(x^*) = g^*(x^*, 0)$.

3.1.6 设 $f : I\!R^n \times I\!R^m \to (-\infty, +\infty]$ 为真凸函数, $A : I\!R^n \to I\!R^m$ 为线性变换. 令

$$g(y) = \inf\{f(x, Ax + y) \mid x \in I\!R^n\}, \quad \forall y \in I\!R^m.$$

试证: (1) g 是 $I\!R^m$ 上的凸函数; (2) $g^*(y^*) = f^*(-A^\tau y^*, y^*)$, 其中 A^τ 为 A 的转置.

3.2 凸集的承托函数

在线性规划中遇到的一类问题是寻求 $I\!R^n$ 上线性函数 $\langle \cdot, x^* \rangle$ 在某个凸函数 K 上的极大 (或极小). 一种更有效的办法是研究该极值随 x^* 的变化. 换句话说, 把此极值作为 x^* 的函数来进行研究. 通常我们称

$$\sigma(x^*|K) \triangleq \sup\{\langle x, x^* \rangle \mid x \in K\}, \quad \forall x^* \in I\!R^n,$$

为 K 的承托函数, 这里我们约定 $\sup \emptyset = -\infty$. 承托函数也叫做支撑函数.

特别从定义直接得到, 当 $K = I\!R^n$ 时,

$$\sigma(x^*|I\!R^n) = \begin{cases} 0, & \text{如果 } x^* = 0, \\ +\infty, & \text{如果 } x^* \in I\!R^n \setminus \{0\}. \end{cases}$$

K 的承托函数 $\sigma(\cdot|K)$ 的有效定义域是指集合

$$\operatorname{dom}(\sigma(\cdot|K)) = \{x^* \in I\!R^n \mid \exists \beta \in I\!R \text{使} \langle x, x^* \rangle \leq \beta, \forall x \in K\}.$$

由此容易看出, $\operatorname{dom}(\sigma(\cdot|K))$ 是 $I\!R^n$ 中的一个凸锥, 叫做凸集 K 的障碍锥,

定理 3.2.1 设 K 是 $\mathbb{R}^n$ 中的非空凸集，则

(1) $D \triangleq \mathrm{dom}\,(\sigma(\cdot|K))$ 是 $\mathbb{R}^n$ 中的凸锥；

(2) $\sigma(\cdot|K)$ 是 $\mathbb{R}^n$ 上的真凸锥函数；

(3) $\sigma(\cdot|K)$ 还是正齐性的.

证明 设 $x^* \in D$, 则对于 $\lambda \geq 0$, 我们有

$$\sigma(\lambda x^*|K) = \sup\{\langle x, \lambda x^*\rangle \mid x \in K\} = \lambda\sigma(x^*|K) < \infty,$$

于是 $\lambda x^* \in D$. 同时上式也表明 $\sigma(\cdot|K)$ 的正齐性. 今设 $x^*, y^* \in D$. 于是对于 $\alpha \geq 0, \beta \geq 0, \alpha + \beta = 1$, 我们有

$$\begin{aligned}
\sigma(\alpha x^* + \beta y^*|K) &= \sup_{x\in K}\langle x, \alpha x^* + \beta y^*\rangle \\
&= \sup_{x\in K}\{\alpha\langle x, x^*\rangle + \beta\langle x, y^*\rangle\} \\
&\leq \sup_{x\in K}\{\alpha\langle x, x^*\rangle\} + \sup_{x\in K}\{\beta\langle x, y^*\rangle\} \\
&= \alpha\sigma(x^*|K) + \beta\sigma(y^*|K) < \infty.
\end{aligned}$$

这表明 $\alpha x^* + \beta y^* \in D$, 从而 D 是凸锥. 同时上式也表明函数 $\sigma(\cdot|K)$ 的凸性. ∎

定理 3.2.2 设 $K \subset \mathbb{R}^n$ 为非空紧凸集，并设 $\sigma(\cdot|K)$ 为其承托函数. 设 $y \neq 0$ 是 $\mathbb{R}^n$ 的任一固定点，那么，

(1) 存在一点 $x(y) \in K$, 使得 $\sigma(y|K) = \langle x(y), y\rangle$;

(2) 超平面 $H \triangleq \{x \in \mathbb{R}^n \mid \langle x, y\rangle = \sigma(y|K)\}$ 为 K 的通过点 $x(y)$ 的承托超平面；

(3) H 到 0 点的距离等于 $\sigma(y/\|y\|\,|K)$.

证明 (1) 由内积的连续性和 K 的紧性得到.

(2) 注意对于 $x \in K$,

$$\langle x, y\rangle \leq \sigma(y|K) = \sup\{\langle x', y\rangle \mid x' \in K\},$$

从而 H 在 $x(y)$ 处承托集合 K.

(3) 由于 y 是超平面 H 的法向量，故存在实数 λ 使得 $\lambda y \in H$. 显然 $\|\lambda y\|$ 是 H 到原点 0 的距离. 注意 $\lambda y \in H$ 意味着 $\langle \lambda y, y\rangle = \sigma(y|K)$, 因此

$$\sigma(y/\|y\|\,|K) = \frac{1}{\|y\|}\sigma(y|K) = \frac{\langle \lambda y, y\rangle}{\|y\|} = \|\lambda y\|.$$

∎

下面我们列举凸集承托函数的几个例子.

例 3.2.1 设 S 为 $\mathbb{R}^n$ 中的闭单位球 (见图 3.2.1). 显然 $\mathrm{dom}\,(\sigma(\cdot|S)) = \mathbb{R}^n$. 对于 $y \neq 0$, 我们有

$$\sigma(y|S) = \sup\{\langle x, y\rangle \mid x \in S\} = \langle y/\|y\|, y\rangle = \|y\|.$$

此外， $\sigma(0|S) = 0$, 因此 $\sigma(y|S) = \|y\|$, $\forall y \in \mathbb{R}^n$. 从而单位球的承托函数对应于空间 $\mathbb{R}^n$ 的欧氏范数.

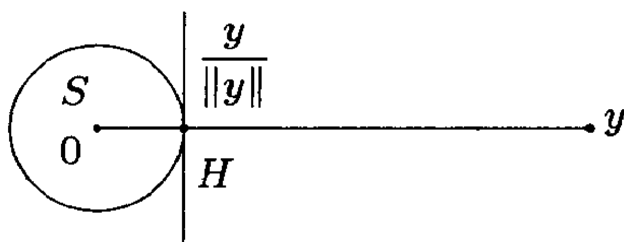


图 3.2.1 闭单位球的承托函数

例3.2.2 设 E 是 $\mathbb{R}^2$ 中顶点在 $(0,\pm1)^\tau$, $(\pm1,0)^\tau$ 的正方形 (见图 3.2.2). 由于 E 有界， $\mathrm{dom}\,(\sigma(\cdot|E)) = \mathbb{R}^n$. 根据定义，对于 $(\xi,\eta)^\tau \in \mathbb{R}^2$,

$$\sigma((\xi,\eta)^\tau|E) = \sup\{\alpha\xi + \beta\eta \mid (\alpha,\beta)^\tau \in E\}.$$

注意 $|\alpha| + |\beta| \le 1$, $\forall(\alpha,\beta)^\tau \in E$, 我们有

$$\sigma((\xi,\eta)^\tau|E) \le \max(|\xi| + |\eta|).$$

另一方面，显然有

$$\begin{aligned}\sigma((\xi,\eta)^\tau|E) &\geq \sup\{\alpha\xi \mid (\alpha,0)^\tau \in E\}\\ &= \sup\{\alpha\xi \mid |\alpha| \leq 1\} = |\xi|.\end{aligned}$$

同理可得 $\sigma((\xi,\eta)^\tau|E) \geq |\eta|$. 因此,

$$\sigma((\xi,\eta)^\tau|E) = \max(|\xi|,|\eta|), \quad \forall (\xi,\eta)^\tau \in \mathbb{R}^2.$$

从定理 1.6.6 可知， $\sigma((\xi,\eta)^\tau|E)$ 中的上确界在 E 的一个极点处达到. 特别，对于 $u=(2,1)^\tau$, 上确界在顶点 $(1,0)^\tau$ 处达到，其值为

$$\sigma(u|E) = 2.$$

如果取 $v=(-1,-2)^\tau$, 则相应的承托函数超平面 H_v 通过点 $(0,-1)^\tau$, 并且 $\sigma(v|E)=1$.

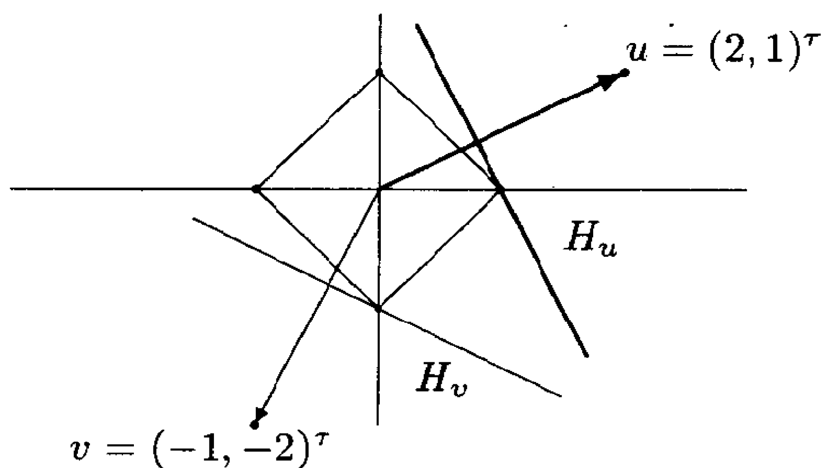


图 3.2.2　正方形情形

例 3.2.3　设 S 是 $\mathbb{R}^2$ 中的无界区域

$$S = \{(\xi,\eta)^\tau \in \mathbb{R}^2 \mid 0 \leq \eta \leq \xi - 1\}.$$

(见图 3.2.3) 由于 S 无界， $D = \operatorname{dom}(\sigma(\cdot|S))$ 是 $\mathbb{R}^2$ 的一个真子集，并且

$$D = \{(\xi,\eta)^{\tau} \mid \xi \le 0, \xi+\eta \le 0\},$$
$$\sigma((\xi,\eta)^{\tau}|S) = \xi, \quad \forall\, (\xi,\eta)^{\tau} \in D.$$

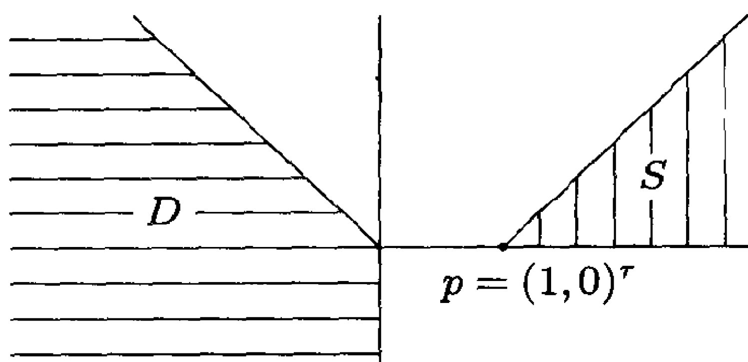


图 3.2.3 无界集的承托函数

例 3.2.4 设 K 是 $I\!R^2$ 中的无界区域.

$$K = \{(\xi,\eta)^{\tau} \mid \xi < 0,\ \eta \le \xi^{-1}\}.$$

(见图 3.2.4) 容易看出 $D = \operatorname{dom}(\sigma(\cdot|K))$ 是 $I\!R^2$ 的非负象限，并且，

$$\sigma((\xi,\eta)^{\tau}|K) = \begin{cases} -2(\xi\eta)^{1/2}, & \forall \xi \ge 0,\ \eta \ge 0, \\ +\infty & \text{其余处}. \end{cases}$$

例 3.2.5 考虑例 3.2.2 推广到 $I\!R^n$. 取

$$E = \{x = (\xi_1,\cdots,\xi_n)^{\tau} \mid |\xi_1| + \cdots + |\xi_n| \le 1\},$$

则对于 $x^* = (\xi_1^*,\cdots,\xi_n^*)^{\tau} \in I\!R^n$,

$$\sigma(x^*|E) = \max\{|\xi_j^*| \mid j = 1,\cdots,n\}.$$

例 3.2.6 设 S 为 $I\!R^n$ 中的凸子集

$$S = \{x = (\xi_1,\cdots,\xi_n)^{\tau} \mid \xi_j \ge 0,\ \xi_1 + \cdots + \xi_n = 1\},$$

则对于 $x^* = (\xi_1^*, \cdots, \xi_n^*)^\tau \in I\!R^n$.

$$\sigma(x^*|S) = \max\{\xi_j^* \mid j = 1, \cdots, n\}.$$

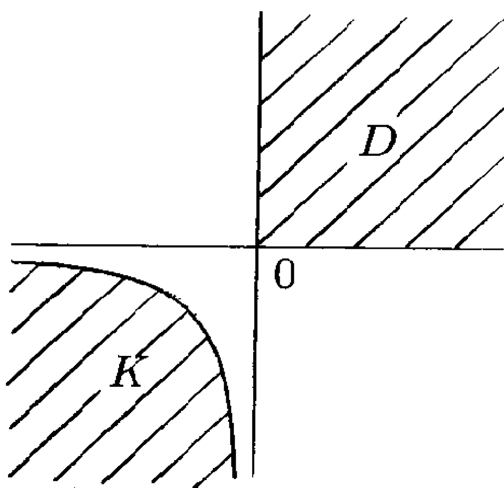


图 3.2.4

定理 3.2.3 设 $K \subset I\!R^n$ 为凸子集，那么其承托函数 $\sigma(\cdot|K)$ 满足

$$\sigma(x^*|K) = \sigma(x^*|\operatorname{cl} K) = \sigma(x^*|\operatorname{ri} K), \quad \forall x^* \in I\!R^n. \tag{3.2.1}$$

证明 从定义直接推出. ▌

定理 3.2.4 设 $K \subset I\!R^n$ 为非空凸子集，那么

(1) $x \in \operatorname{cl} K$ 等价于

$$\langle x, x^* \rangle \le \sigma(x^*|K), \quad \forall x^* \in I\!R^n; \tag{3.2.2}$$

(2) $x \in \operatorname{ri} K$ 等价于

$$\begin{cases} \langle x, x^* \rangle \le \sigma(x^*|K), & \forall x^* \in I\!R^n, \\ \langle x, x^* \rangle < \sigma(x^*|K), & \forall x^* \in S(K), \end{cases} \tag{3.2.3}$$

其中 $S(K) = \{x^* \in I\!R^n \mid \sigma(x^*|K) \neq -\sigma(-x^*|K)\}$;

(3) $x \in \operatorname{int} K$ 等价于

$$\langle x, x^* \rangle < \sigma(x^*|K), \quad \forall x^* \in \mathbb{R}^n \backslash \{0\}; \tag{3.2.4}$$

(4) $x \in \operatorname{aff} K$ 等价于

$$\langle x, x^* \rangle = \sigma(x^*|K), \quad \forall x^* \in M(K), \tag{3.2.5}$$

其中 $M(K) = \{x^* \in \mathbb{R}^n \mid \sigma(x^*|K) = -\sigma(-x^*|K)\}$.

证明 (1) 只须注意由定理 1.5.8, 闭凸集 $\operatorname{cl} K$ 是一切包含 K 的闭半空间之交.

(2) 设 $x_0 \in \operatorname{ri} K$, 则由 (1) 得知式 (3.2.2) 成立. 今若 $\sigma(x^*|K) \neq -\sigma(-x^*|K)$, 则 K 不可能包含在形如 $H(x^*, \alpha) = \{x \in \mathbb{R}^n \mid \langle x^*, x \rangle = \alpha\}$ 的超平面内. 从定理 1.6.1 可知 $\langle x_0, x^* \rangle < \sigma(x^*|K)$ 成立, 因为否则的话, 通过 $x_0 \in \operatorname{ri} K$ 可以作 K 的承托超平面了! 反之设 x_0 满足式 (3.2.3). 如果 $x_0 \notin \operatorname{ri} K$, 则通过 x_0 可以作 K 的承托超平面 $H = \{x \in \mathbb{R}^n \mid \langle x^*, x \rangle = \langle x^*, x_0 \rangle\}$. 由于 $K \not\subset H$, 故必有

$$\sigma(x^*|K) \neq -\sigma(-x^*|K).$$

于是 $\langle x_0, x^* \rangle < \sigma(x^*|K) = \langle x_0, x^* \rangle$, 得出矛盾, 所以 $x_0 \in \operatorname{ri} K$.

(3) 注意当 $\operatorname{int} K \neq \emptyset$ 时, K 不可能包含在任何超平面内, 故

$$\sigma(x^*|K) \neq -\sigma(-x^*|K), \quad \forall x^* \in \mathbb{R}^n \backslash \{0\}.$$

于是从 (2) 得知 (3) 成立.

(4) 只须注意包含 K 的最小仿射集 $\operatorname{aff} K$ 恰好是包含 K 的一切超平面之交. ▮

推论 3.2.5 设 M_1 和 M_2 为 $\mathbb{R}^n$ 的凸集, 那么,

$$\operatorname{cl} M_1 \subset \operatorname{cl} M_2 \Longleftrightarrow \sigma(\cdot|M_1) \leq \sigma(\cdot|M_2).$$

由此可见，$\mathbb{R}^n$ 中的闭凸集 K 可以表示成由其承托函数描述的一组不等式的解集：

$$K = \{x \in \mathbb{R}^n \mid \langle x, x^* \rangle \leq \sigma(x^*|K),\ \forall\, x^* \in \mathbb{R}^n\}.$$

这样，K 完全由其承托函数决定. 这个事实当然是很有意义的，因为它表明，$\mathbb{R}^n$ 中闭凸集与 $\mathbb{R}^n$ 上的某些函数之间有着一一对应关系.

这种对应关系具有许多重要性质. 例如，两个非空凸集 K_1 和 K_2 之和的承托函数对应于两个凸集 K_1 和 K_2 之承托函数之和 (注意凸集 $K_1 + K_2$ 的定义):

$$\begin{aligned}
&\sigma(x^*|K_1 + K_2) \\
&= \sup\{\langle x_1 + x_2, x^* \rangle \mid x_1 \in K_1,\ x_2 \in K_2\} \\
&= \sup\{\langle x_1, x^* \rangle \mid x_1 \in K_1\} + \sup\{\langle x_2, x^* \rangle \mid x_2 \in K_2\} \\
&= \sigma(x^*|K_1) + \sigma(x^*|K_2).
\end{aligned}$$

同样地，对于非空凸集 $K \subset \mathbb{R}^n$ 和正实数 λ, 我们有

$$\begin{aligned}
\sigma(x^*|\lambda K) &= \sup\{\langle x, x^* \rangle \mid x \in \lambda K\} \\
&= \sup\{\langle \lambda x^*, x/\lambda \rangle \mid x/\lambda \in K\} \\
&= \lambda \sigma(x^*|K).
\end{aligned}$$

因此集合的加法变成函数的加法，而集合与正常数相乘变成函数与正常数相乘.

那么到底 $\mathbb{R}^n$ 上什么样的函数才是一个凸集的承托函数呢？换句话说，给定 $\mathbb{R}^n$ 上的一个函数 f, 如何来判断

它是否是一个集合 K 的承托函数？下面我们来回答这个问题.

首先注意，凸集 K 与示性函数 $\delta(\cdot|K)$ 之间有着一一对应关系. 依照定义， $\delta(\cdot|K)$ 的共轭函数是

$$\sup_{x\in \mathbb{R}^n}\{\langle x,x^*\rangle-\delta(x|K)\}=\sup_{x\in K}\langle x,x^*\rangle,$$

即

$$(\delta(\cdot|K))^*(x^*)=\sigma(x^*|K),\quad \forall\, x^*\in \mathbb{R}^n.$$

由此，并根据定理 3.1.2, $\sigma(\cdot|K)$ 的共轭是

$$(\sigma(\cdot|K))^*(x)=\mathrm{cl}\,\delta(x|K)=\delta(x|\mathrm{cl}\,K),\quad \forall\, x\in \mathbb{R}^n.$$

这样我们证明了如下定理：

定理 3.2.6 $\mathbb{R}^n$ 中闭凸集的示性函数和承托函数是互为共轭的.

定理 3.2.7 为了函数 $f:\mathbb{R}^n\to(-\infty,+\infty]$ 是 $\mathbb{R}^n$ 的某个非空闭凸集的承托函数，必须且只须它是正齐性的闭真凸函数.

证明 必要性根据定理 3.2.1 证得. 现在证明充分性. 设 $\lambda>0$, 则利用 f 的正齐性有

$$\begin{aligned}\lambda f^*(x^*)&=\lambda\sup\{\langle x,x^*\rangle-f(x)\mid x\in \mathbb{R}^n\}\\&=\sup\{\langle \lambda x,x^*\rangle-f(\lambda x)\mid x\in \mathbb{R}^n\}\\&=\sup\{\langle x,x^*\rangle-f(x)\mid x\in \mathbb{R}^n\}\\&=f^*(x^*),\quad \forall\, x^*\in \mathbb{R}^n.\end{aligned}$$

这表明 f^* 在 $\mathbb{R}^n$ 上只有可能取值 0 或 $+\infty$, 即 f^* 是集合 $K=\{x^*\in \mathbb{R}^n\mid f^*(x^*)=0\}$ 的示性函数. 由于 f^* 是闭凸函数，故 K 是闭凸集. 因此 $f=f^{**}=\sigma(\cdot|K)$. ∎

特别地，定理 3.2.7 表明 $\sigma(\cdot|K)$ 是 $\mathbb{R}^n$ 上的下半连续函数，并且

$$\sigma(x_1^*+x_2^*|K)\le\sigma(x_1^*|K)+\sigma(x_2^*|K),\quad \forall\, x_1^*,x_2^*\in\mathbb{R}^n.$$

推论 3.2.8 设 $f:\mathbb{R}^n\to(-\infty,+\infty]$ 是一个正齐性的真凸函数，那么 $\mathrm{cl}\, f$ 是某个闭凸集 $K\subset\mathbb{R}^n$ 的承托函数，并且

$$K=\{x^*\in\mathbb{R}^n\mid\langle x,x^*\rangle\le f(x),\ \forall\, x\in\mathbb{R}^n\}.$$

证明 我们知道，$\mathrm{cl}\, f$ 是一个正齐性的闭真凸函数．于是存在某个闭凸集 $K\subset\mathbb{R}^n$，使得 $f^*=(\mathrm{cl}\, f)^*=\delta(\cdot|K)$，并且 $K=\{x^*\in\mathbb{R}^n\mid f^*(x^*)\le 0\}$．但是 $f^*(x^*)\le 0$ 等价于 $\langle x,x^*\rangle-f(x)\le 0$, $\forall x\in\mathbb{R}^n$. ∎

推论 3.2.9 $\mathbb{R}^n$ 中非空有界凸函数的承托函数是有穷值正齐性的凸函数.

定理 3.2.10 设 f 是 $\mathbb{R}^n$ 上闭的真凸函数．那么凸集 $\{x\in\mathbb{R}^n\mid f(x)\le 0\}$ 的承托函数是 $\mathrm{cl}\, g$，其中 g 是由 f 生成的正齐性凸函数．同样地，由 f 生成的正齐性凸函数的闭包则是 $\{x^*\in\mathbb{R}^n\mid f^*(x^*)\le 0\}$ 集合的承托函数.

证明 由于 f 和 f^* 之间的对偶关系，只须证明第一个结论．依照推论 3.2.8, $\mathrm{cl}\, h$ 是集合

$$D=\{x^*\in\mathbb{R}^n\mid\langle x,x^*\rangle\le h(x),\ \forall\, x\in\mathbb{R}^n\}$$

的承托函数．但依 h 的定义，

$$h(x)=\inf\{(f\lambda)(x)\mid\lambda>0\},\ \forall x\in\mathbb{R}^n,$$

其中 $(f\lambda)(x)=\frac{1}{\lambda}f(\lambda x)$, $\lambda>0$．由此可见，对于 $x^*\in\mathbb{R}^n$,

$$\langle x,x^*\rangle\le h(x),\ \forall\, x\in\mathbb{R}^n\iff\langle x,x^*\rangle\le f(x),\ \forall x\in\mathbb{R}^n.$$

因此

$$D = \{x^* \in I\!R^n \mid \langle x, x^* \rangle \le f(x),\ \forall\, x \in I\!R^n\}$$
$$= \{x^* \in I\!R^n \mid f^*(x^*) \le 0\}.$$

∎

作为例子，我们来计算"椭圆型"凸集

$$M = \left\{x \in I\!R^n \mid \frac{1}{2}\langle Qx, x\rangle + \langle a, x\rangle + \alpha \le 0\right\}$$

的承托函数，这里 Q 是 $n \times n$ 正定对称矩阵. 记

$$f(x) = \frac{1}{2}\langle Qx, x\rangle + \langle a, x\rangle + \alpha,$$

则 $M = \{x \in I\!R^n \mid f(x) \le 0\}$, f 是 $I\!R^n$ 上有穷值凸函数. 依定理 3.2.10, M 的承托函数 $\sigma(\cdot \mid M)$ 是由 f^* 所生成的正齐性凸函数 g 的闭包. 上一节中我们已经知道，

$$f^*(x^*) = \frac{1}{2}\Big\langle x^* - a, Q^{-1}(x^* - a)\Big\rangle - \alpha$$
$$= \frac{1}{2}\Big\langle x^*, Q^{-1}x^*\Big\rangle + \langle b, x^*\rangle + \beta,$$

其中 $b = -Q^{-1}a$, $\beta = \frac{1}{2}\langle a, Q^{-1}a\rangle - \alpha$. 从 2.3 节我们知道，对于 $x^* \ne 0$, 有

$$g(x^*) = \inf\{(f^*\lambda)(x^*) \mid \lambda > 0\}$$
$$= \inf\{\lambda f^*(\lambda^{-1}x^*) \mid \lambda > 0\}.$$

由于 $\operatorname{dom} f^* = I\!R^n$, 我们有 $\operatorname{dom} g = I\!R^n$. 因此 g 本身是闭的，并且

$$\sigma(x^*|M) = g(x^*) = \inf_{\lambda>0}\{(1/2\lambda)\langle x^*, Q^{-1}x^*\rangle + \langle b, x^*\rangle + \lambda\beta\}.$$

假定 $M \ne \emptyset$, 我们有

$$0 \le \sup\{-f(x) \mid x \in I\!R^n\} = f^*(0) = \beta.$$

如果 $\beta = 0$, 则 $g(x^*)$ 中的下确界显然是 $\langle b, x^*\rangle$. 如果 $\beta > 0$,

则通过对 λ 求导可算出此下确界：

$$\sigma(x^*|M) = \langle b, x^* \rangle + [2\beta\langle x^*, Q^{-1}x^* \rangle]^{1/2}.$$

定理 3.2.11 设 f 是 $\mathbb{R}^n$ 上的真凸函数. 那么 $\operatorname{dom} f$ 的承托函数是 f^* 的回收函数 f^*0^+. 如果 f 还是闭的，则 $\operatorname{dom} f^*$ 的承托函数是 f 的回收函数 $f0^+$.

证明 根据定义，$f^*(x^*)$ 是仿射函数

$$h(x^*) = \langle x, x^* \rangle - \mu$$

对于一切 $(x, \mu) \in \operatorname{epi} f$ 的上确界. 因此，$\operatorname{epi} f^*$ 是相应的闭半空间 $\operatorname{epi} h$ 之 (非空) 交. 于是根据推论 2.3.8, 回收锥 $0^+(\operatorname{epi} f^*)$ 是相应集合 $0^+(\operatorname{epi} h)$ 之交. 这表明 f^*0^+ 是诸函数 $h0^+$ 的点点上确界. 但显然，当 $h(x^*) = \langle x, x^* \rangle - \mu$ 时，$(h0^+)(x^*) = \langle x, x^* \rangle$. 所以 $(f^*0^+)(x^*)$ 是线性函数 $\langle x, x^* \rangle$ 关于 $(x, \mu) \in \operatorname{epi} f$ 的上确界，即

$$\begin{aligned}(f^*0^+)(x^*) &= \sup\{\langle x, x^* \rangle \mid x \in \operatorname{dom} f\} \\ &= \sigma(x^*|\operatorname{dom} f).\end{aligned}$$

定理的第二个结论从对偶性推出，因为当 f 闭真凸函数时，$f^{**} = f$. ∎

习 题 3.2

3.2.1 设 f 是凸锥 $K \subset \mathbb{R}^n$ 上的实值函数，假定 f 是正齐性凸函数，试证：f 是次加法的，即 $f(x+y) \leq f(x) + f(y)$, $\forall x, y \in K$.

3.2.2 设 $S = \{(\xi_1, \cdots, \xi_n)^\tau \in \mathbb{R}^n | \xi_i \geq 0, \xi_1 + \cdots + \xi_n = 1\}$, 试证：对于 $y = (\eta_1, \cdots, \eta_n)^\tau \in \mathbb{R}^n$,

$$\sigma(y|S) = \max\{\eta_i \mid i = 1, 2, \cdots, n\}.$$

3.2.3 设 $S = \{(\xi_1, \cdots, \xi_n)^\tau \in \mathbb{R}^n \mid |\xi_1| + \cdots + |\xi_n| \leq 1\}$, 试证：对于 $y = (\eta_1, \cdots, \eta_n)^\tau \in \mathbb{R}^n$,

$$\sigma(y|S) = \max\{|\eta_i| \mid i = 1, 2, \cdots, n\}.$$

3.2.4 设 $\overline{B}(0,1)$ 为 $\mathbb{R}^n$ 中闭单位球, $a \in \mathbb{R}^n$, $\lambda > 0$. 试找出集合 $a + \lambda\overline{B}(0,1)$ 的承托函数.

3.2.5 设 $S = \{(\xi_1, \xi_2)^\tau \in \mathbb{R}^2 \mid \xi_1 < 0,\ \xi_1\xi_2 \le 1\}$, 试证:

$$\sigma(y|S) = \begin{cases} -2(\eta_1\eta_2)^{1/2}, & \text{如果 } y = (\eta_1, \eta_2)^\tau \ge 0, \\ +\infty, & \text{其余处}. \end{cases}$$

3.2.6 设 $S = \{(\xi_1, \xi_2)^\tau \in \mathbb{R}^2 \mid 2\xi_1 + \xi_2^2 \le 0\}$, 试证:

$$\sigma(y|S) = \begin{cases} \eta_2^2/2\eta_1, & \text{如果 } \eta_1 > 0, \\ 0, & \text{如果 } \eta_1 = 0 = \eta_2, \\ +\infty, & \text{其余处}. \end{cases}$$

3.2.7 设 M_1 和 M_2 是 $\mathbb{R}^n$ 的两个非空凸子集, 试证:

$$\sigma(x|M_1 + M_2) = \sigma(x|M_1) + \sigma(x|M_2), \quad \forall\, x \in \mathbb{R}^n,$$

$$\mathrm{dom}\,(\sigma(\cdot|M_1 + M_2)) = \mathrm{dom}\,(\sigma(\cdot|M_1)) \cap \mathrm{dom}\,(\sigma(\cdot|M_2)).$$

3.2.8 设 M_1 和 M_2 是 $\mathbb{R}^n$ 的非空紧凸集, 试证:

$$M_1 \subset M_2 \Longleftrightarrow \sigma(x|M_1) \le \sigma(x|M_2), \quad \forall\, x \in \mathbb{R}^n.$$

3.2.9 设 A, B 为 $\mathbb{R}^n$ 的紧凸集, $0 \in \mathrm{int}\, A \cap \mathrm{int}\, B$. 试证:

(1) $\sigma(x|a + A) = \sigma(x|A) + \langle a, x\rangle$, $\forall\, x \in \mathbb{R}^n, a \in \mathbb{R}^n$;

(2) $\sigma(x|\mathrm{conv}\,(A \cup B)) = \max\{\sigma(x|A), \sigma(x|B)\}$, $\forall\, x \in \mathbb{R}^n$;

(3) $p_{A\cap B}(x) = \max\{p_A(x), p_B(x)\}$, $\forall x \in \mathbb{R}^n$, 其中 p_M 为 M 的 Minkowski 函数.

3.2.10 设 p 是定义在 $\mathbb{R}^n$ 上的实值函数, 满足

$$p(x + y) \le p(x) + p(y), \quad \forall\, x, y \in \mathbb{R}^n,$$

$$p(\alpha x) = |\alpha| p(x), \quad \forall\, x \in \mathbb{R}^n, \alpha \in R.$$

(这样的函数称为 $\mathbb{R}^n$ 上的半范数) 求证:

(1) p 是凸函数；

(2) $p(0)=0$;

(3) $p(x)\geq 0,\ \forall x\in \mathbb{R}^n$;

(4) $|p(x)-p(y)|\leq p(x-y),\ \forall x,y\in \mathbb{R}^n$.

3.2.11 完成推论 3.2.9 的证明.

3.3 极化集

凸集和其承托函数之间的对应关系反映了凸函数的正齐性和示性函数之间的某种对偶性，即对于 $\mathbb{R}^n$ 上的一个真凸函数 f, 如果 f 是正齐性的，则其共轭 f^* 是某个凸集的示性函数；反之，如果 f 是某个凸集的示性函数，则 f^* 是正齐性的. 由此可见，如果 f 是一个正齐性的示性函数，则 f^* 也是一个正齐性的示性函数. 显然正齐性的示性函数恰好是某个锥的示性函数. 于是，若 $f(x)=\delta(x|K)$, 其中 K 为 $\mathbb{R}^n$ 的一个非空凸锥，则 $f^*(x^*)=\delta(x^*|K^o)$, 其中 K^o 为另一个非空闭凸锥，叫做 K 的极化锥或负对偶锥 [1]. 根据推论 3.2.8,

$$\begin{aligned}K^o &= \{x^*\in \mathbb{R}^n \mid \langle x,x^*\rangle \leq \delta(x|K),\forall x\in \mathbb{R}^n\}\\ &= \{x^*\in \mathbb{R}^n \mid \langle x,x^*\rangle \leq 0,\forall x\in K\}.\end{aligned} \tag{3.3.1}$$

K^o 作为凸锥又有其自身的极化锥 $K^{oo}=(K^o)^o=\operatorname{cl}K$. 此外，不难看出， $K^o=(\operatorname{cl}K)^o$. 这样，凸函数之间的共轭对应关系也反应了凸锥之间的一种特殊的一一对应关系：

定理 3.3.1 设 $K\subset \mathbb{R}^n$ 为非空闭凸锥. 那么 K 的极

1) 凸锥 K 的对偶锥 K^* 定义成 $K^*=\{x^*\in \mathbb{R}^n \mid \langle x,x^*\rangle \geq 0,\forall x\in K\}$.

化锥 K^o 也是一个闭凸锥，并且 $K^{oo} = K$. K 和 K^o 的示性函数互为共轭.

极化锥的概念也可以推广到 $\mathbb{R}^n$ 的任意凸集 M(甚至任意非空集). 从 2.3 节可知， M 的度规函数 $\gamma(\cdot|M)$ 是由 $f = \delta(\cdot|M) + 1$ 所生成的正齐性凸函数. 根据定理 3.2.10, $\gamma(\cdot|M)$ 的闭包 $\mathrm{cl}\,\gamma(\cdot|M)$ 是集合 $\{x^* \in \mathbb{R}^n \mid f^*(x^*) \le 0\}$ 的承托函数，但是 $f^* = \sigma(\cdot|M) - 1$, 因此存在一个闭凸集 M^o 使得

$$\mathrm{cl}\,\gamma(x^*|M) = \sigma(x^*|M^o),\ \forall\, x^* \in \mathbb{R}^n,$$

并且

$$\begin{aligned} M^o &= \{x^* \in \mathbb{R}^n \mid \sigma(x^*|M) - 1 \le 0\} \\ &= \{x^* \in \mathbb{R}^n \mid \langle x, x^* \rangle \le 1,\ \forall\, x \in M\}. \end{aligned} \tag{3.3.2}$$

M^o 叫做 M 的极化集. 显然 M^o 包含原点. 同样可以定义 M^o 的极化集 $M^{oo} = (M^o)^o$:

$$\begin{aligned} M^{oo} &= \{x^* \in \mathbb{R}^n \mid \langle x, x^* \rangle \le 1,\ \forall\, x \in M^o\} \\ &= \{x \in \mathbb{R}^n \mid \sigma(x|M^o) \le 1\} \\ &= \{x \in \mathbb{R}^n \mid \mathrm{cl}\,\gamma(x|M) \le 1\}. \end{aligned} \tag{3.3.3}$$

在进一步讨论极化集的性质之前，我们先观察几个例子.

例 3.1.1 设 $M = \{a\}$ 为 $\mathbb{R}^n$ 中的单点集, 并且 $a \ne 0$, 那么 M 的极化集是半空间

$$M^o = \{x^* \in \mathbb{R}^n \mid \langle x^*, a \rangle \le 1\};$$

又当 $a = 0$ 时， $M^o = \mathbb{R}^n$, 因为

$$\langle x^*, 0 \rangle = 0 \le 1, \quad \forall\, x^* \in \mathbb{R}^n.$$

例 3.3.2 设 L 是 $\mathbb{R}^n$ 的一子空间，则

$$L^o = \{y \in \mathbb{R}^n \mid \langle y, x\rangle \le 1,\ \forall x \in L\}$$
$$= \{y \in \mathbb{R}^n \mid \langle y, x\rangle = 0,\ \forall x \in L\} = L^{\perp}.$$

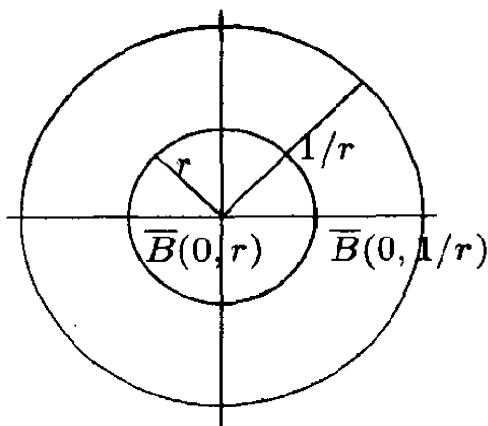


图 3.3.1 球的极化集

例 3.3.3 设 $\overline{B}(0,r)$ 为 $\mathbb{R}^n$ 中的以 0 为中心、r 为半径的闭球．那么

$$[\overline{B}(0,r)]^o = \overline{B}(0,1/r).$$

事实上，如果 $y \in [\overline{B}(0,r)]^o$, 记 $y' = ry/\|y\|$, 则 $\|y'\| = r$, 即 $y' \in \overline{B}(0,r)$. 因此

$$\langle y', y\rangle = \frac{r}{\|y\|}\langle y, y\rangle = r\|y\| \le 1,$$

这表明 $\|y\| \le 1/r$, 故 $y \in \overline{B}(0,1/r)$ (见图 3.3.1). 反之，若 $z \in \overline{B}(0,1/r)$, 则对任意 $x \in \overline{B}(0,r)$, 我们有

$$\langle x, z\rangle \le \|x\|\,\|z\| \le r(1/r) = 1,$$

这表明 $z \in [\overline{B}(0,r)]^o$. 从而 $[\overline{B}(0,r)]^o = \overline{B}(0,1/r)$.

对于任意非空集 $M \subset \mathbb{R}^n$, 我们也可以定义 M 的极化集 M^o:

$$M^o = \{x^* \in \mathbb{R}^n \mid \langle x, x^* \rangle \le 1,\ \forall\, x \in M\}. \tag{3.3.4}$$

当 M 是凸锥 K 时，这里的极化集与式 (3.3.1) 中极化锥 K^o 的定义是相同的.

定理 3.3.2 设 M, M_1, M_2 和 $M_i (i \in I)$ 为 $\mathbb{R}^n$ 的非空集. 那么,

(1) $\left[\bigcup_{i \in I} M_i\right]^o = \bigcap_{i \in I} M_i^o$;

(2) M^o 是含原点 0 的闭凸集;

(3) $M_1 \subset M_2 \Longrightarrow M_2^o \subset M_1^o$;

(4) $\lambda > 0 \Longrightarrow (\lambda M)^o = (1/\lambda) M^o$.

证明 (1) 从极化集定义式 (3.3.4), 我们有

$$\begin{aligned}\left[\bigcup_{i \in I} M_i\right]^o &= \{x^* \mid \langle x, x^* \rangle \le 1,\ \forall\, x \in M_i, \forall i \in I\} \\ &= \bigcap_{i \in I} \{x^* \mid \langle x, x^* \rangle \le 1,\ \forall\, x \in M_i\} \\ &= \bigcap_{i \in I} M_i^o.\end{aligned}$$

(2) 容易直接从定义验证.

(3) 从 $M_1 \subset M_2$ 知 $M_2 = M_1 \cup (M_2 \backslash M_1)$. 于是从已经证明的 (1) 得到 $M_2^o = M_1^o \cap (M_2 \backslash M_1)^o \subset M_1^o$.

(4) 只须注意,

$$\langle x, x^* \rangle \le 1,\ \forall\, x \in \lambda M \Longleftrightarrow \langle x, x^*/\lambda \rangle \le 1,\ \forall x \in M. \quad ▮$$

定理 3.3.3 设 M 为 $\mathbb{R}^n$ 中非空集, 那么

$$M^{oo} = \mathrm{cl}\,\mathrm{conv}\,(M \cup \{0\}).$$

特别当 M 为含原点的闭凸集时, $M^{oo} = M$.

证明 如果记 $M_1 = \mathrm{cl}\,\mathrm{conv}\,(M \cup \{0\})$, 则从定义直接可知 $M_1 \subset M^{oo}$. 如果存在某个 $x_0 \in M^{oo} \backslash M_1$, 则根据凸

集分离的推论 2.2.7, 存在一超平面

$$H = \{x \in \mathbb{R}^n \mid \langle x, x^* \rangle = 1\}$$

把 x_0 与 M_1 严格分离. 由于 $0 \in M_1$, 我们有 $\langle x, x^* \rangle \le 1$, $\forall x \in M_1$; 而 $\langle x_0, x^* \rangle > 1$. 前面的不等式表明 $x^* \in M_1^o \subset M^o$, 于是后面的不等式意味着 $x_0 \notin M^{oo}$, 得出矛盾. 因此 $M^{oo} = M_1$. ▮

综上所述, 我们现在得到一种新的对偶关系, 表成下一定理的形式.

定理 3.3.4 设 $M \subset \mathbb{R}^n$ 为含原点的闭凸集, 则极化集 M^o 为另一个含原点的闭凸集, 并且成立 $M^{oo} = M$. 于是 M 的度规函数 $\gamma(\cdot|M)$ 是 M^o 的承托函数 $\sigma(\cdot|M^o)$, 即 $\gamma(\cdot|M) = \sigma(\cdot|M^o)$, 并且

$$M = \{x \in \mathbb{R}^n \mid \gamma(x|M) \le 1\};$$

而 M^o 的度规函数 $\gamma(\cdot|M^o)$ 是 M 的承托函数, 即 $\gamma(\cdot|M^o) = \sigma(\cdot|M)$.

引理 3.3.5 设 M 为 $\mathbb{R}^n$ 中含原点的闭凸集. 那么 M^o 有界的充分必要条件是 $0 \in \operatorname{int} M$; 同样地, M 有界当且仅当 $0 \in \operatorname{int} M^o$.

证明 从推论 3.2.9 可知, M^o 有界等价于 M^o 的承托函数 $\gamma(\cdot|M)$ 处处有穷. 另一方面, 根据推论 1.4.12, $\gamma(\cdot|M)$ 处处有穷又等价于 $0 \in \operatorname{int} M$. ▮

下面我们进一步讨论多面体. 先定义有关集合的几个概念.

设 M 是 $\mathbb{R}^n$ 的一个集合,

(1) M 叫做吸收集, 是指对于任意 $x \in \mathbb{R}^n$, 存在常数 $\alpha > 0$ 使得 $x \in \beta M, \forall \beta \ge \alpha$.

(2) M 叫做平衡集，是指 $\alpha M \subset M, \forall \alpha \in [-1,1]$.

注意，当 $\operatorname{aff} M \neq \mathbb{R}^n$，即 $\dim(\operatorname{aff} M) < n$ 时，M 不可能是吸收的. 因此对于一个吸收集 M, 必然有 $\operatorname{aff} M = \mathbb{R}^n$.

引理 3.3.6 设 $M \subset \mathbb{R}^n$ 是凸集，则为了 M 是吸收的，必须且只须 $0 \in \operatorname{int} M$.

证明 当 $0 \in \operatorname{int} M$ 时，显然 M 是吸收的 (这时不要求 M 是凸的). 现在设 M 是吸收的凸集，但 $0 \notin \operatorname{int} M$. 于是通过原点 0 可以作一超平面使得 M 位于该超平面的一侧，即存在向量 $z \in \mathbb{R}^n$ 使

$$\langle x, z \rangle \leq 0, \quad \forall x \in M.$$

这样，当向量 $y \in \mathbb{R}^n$ 满足 $\langle y, z \rangle > 0$ 时，$\langle \alpha y, z \rangle > 0$, $\forall \alpha > 0$, 从而 $y \notin \beta M, \forall \beta > 0$. 这表明 M 不是吸收的. ∎

对于包含原点的闭凸集 $M \subset \mathbb{R}^n$, 我们在 2.2 节中定义了的 M 度规函数

$$\gamma(x|M) = \inf\{\alpha \geq 0 \mid x \in \alpha M\}, \quad \forall x \in \mathbb{R}^n.$$

这里为了记号简单起见，有时我们也记 $\gamma_M = \gamma(\cdot|M)$. 我们也指出 γ_M 是 $\mathbb{R}^n$ 上的正齐性次加法凸函数.

引理 3.3.7 设 $M \subset \mathbb{R}^n$ 为包含原点 0 的闭凸集，那么，

(1) 若 M 有界，则 $\gamma_M(x) = 0 \Longleftrightarrow x = 0$;

(2) 若 M 是吸收的，则 $\gamma_M(x) < \infty, \forall x \in \mathbb{R}^n$;

(3) 若 M 是有界吸收且平衡的，则 γ_M 是 $\mathbb{R}^n$ 上的一个范数，即它满足

1) $\gamma_M(x) = 0 \Longleftrightarrow x = 0$,

2) $\gamma_M(\alpha x) = |\alpha| \gamma_M(x), \quad \forall x \in \mathbb{R}^n, \alpha \in \mathbb{R}$,

3) $\gamma_M(x+y) \le \gamma_M(x) + \gamma_M(y), \forall x, y \in I\!R^n$;
(见 1.2 节中欧氏范数的性质)

(4) $M = \{x \in I\!R^n \mid \gamma_M(x) \le 1\}$;

(5) 若 M 是吸收的，则 γ_M 是 $I\!R^n$ 上的连续函数.

(6) 若 $0 \in \operatorname{int} M$, 则

$$\operatorname{int} M = \{x \in I\!R^n \mid \gamma_M(x) < 1\},$$

$$\operatorname{bd} M = \{x \in I\!R^n \mid \gamma_M(x) = 1\}.$$

证明 (1) 和 (2) 可直接从 γ_M 的定义推出. 从第一章我们已经知道 γ_M 的正齐性和次加法性，从 M 均衡性可导出 γ_M 的对称性，即 $\gamma_M(-x) = \gamma_M(x), \forall x \in I\!R^n$. 于是再由 (1) 和 (2) 即得 (3) 成立.

现在证 (4). 当 $x \in M$ 时，显然 $\gamma_M(x) \le 1$. 反之设 $\gamma_M \le 1$. 如果 $\gamma_M(x) < 1$, 则由 $0 \in M$ 和 M 的凸性可知 $x \in M$; 如果 $\gamma_M(x) = 1$, 则可取 $\alpha_n > 1$ 使得 $\alpha_n \to 1 (n \to \infty)$, $x/\alpha_n \in M, \forall n \ge 1$. 但 $x/\alpha_n \to x (n \to \infty)$, 故从 M 的闭性推出 $x \in M$.

当 M 为吸收集时，从 (2) 知道，γ_M 是 $I\!R^n$ 上处处取有穷值的凸函数，因此依推论 2.5.2, γ_M 是 $I\!R^n$ 上的连续函数，即 (5) 成立.

最后证明 (6). 如果 $x \in \operatorname{int} M$, 则必存在 $\epsilon > 0$, 使得

$$(1+\epsilon)x \in M, \text{ 即 } x \in \frac{1}{1+\epsilon} M,$$

因此 $\gamma_M(x) \le \frac{1}{1+\epsilon} < 1$. 反之设 $\gamma_M(x) < 1$, 则从 (5) 中证明的 γ_M 的连续性可知存在 $\epsilon > 0$ 和 $r \in (\gamma_M(x), 1)$ 使得

$$z \in B(x, \epsilon) \Longrightarrow \gamma_M(z) < r < 1,$$

从而 $B(x,\epsilon) \subset M$. 这正是说 $x \in \operatorname{int} M$. 最后一个结论 $\operatorname{bd} M = \{x \in \mathbb{R}^n \mid \gamma_M(x) = 1\}$ 是自然的了. ▌

对于任意集合 $M \subset \mathbb{R}^n$, M^o 总是 $\mathbb{R}^n$ 中含原点的闭凸集，从而可以定义 M^o 的度规函数 $\gamma_{M^o} = \gamma(\cdot|M^o)$.

定理 3.3.8 设 M 是 $\mathbb{R}^n$ 中含原点 0 的闭凸集，那么对于一切 $x \in \mathbb{R}^n$,

$$\gamma_{M^o}(x) = \inf\{\alpha \geq 0 \mid \langle x, y\rangle \leq \alpha\gamma_M(y),\ \forall y \in M\}.$$

证明 根据度规函数的定义,

$$\begin{aligned}\gamma_{M^o}(x) &= \inf\{\alpha > 0 \mid x \in \alpha M^o\}\\ &= \inf\{\alpha > 0 \mid \langle x, y\rangle \leq \alpha,\ \forall y \in M\}.\end{aligned}$$

另一方面，从定理 3.3.4 可知 $\gamma_M(x) = \sigma(x|M^o)$. 因此

$$\langle x, y\rangle \leq \alpha\gamma_M(y), \quad \forall y \in M$$

等价于

$$\langle x, y\rangle \leq \alpha\sigma(y|M^o), \quad \forall y \in M.$$

而由定理 3.2.4 又可知上式等价于 $x/\alpha \in M^o$, 即 $x \in \alpha M^o$. ▌

现在考虑多面体的极化集.

定理 3.3.9 设 $P = \operatorname{conv}\{x_1, \cdots, x_m\}$ 是 $\mathbb{R}^n$ 中的多面体，那么

$$P^o = \bigcap_{i=1}^m \{y \in \mathbb{R}^n \mid \langle y, x_i\rangle \leq 1\} = \bigcap_{i=1}^m H^-(x_i 1),$$

其中 $H^-(x_i, 1) = \{y \in \mathbb{R}^n \mid \langle y, x_i\rangle \leq 1\}$.

证明 根据极化集的定义,

$$\begin{aligned}P^o &= \{y \in \mathbb{R}^n \mid \langle y, x\rangle \leq 1,\ \forall x \in P\}\\ &\subset \{y \in \mathbb{R}^n \mid \langle y, x_i\rangle \leq 1,\ i = 1, \cdots, m\}\\ &= \bigcap_{i=1}^m H^-(x_i, 1).\end{aligned}$$

另一方面，若 $\langle y, x_i\rangle \le 1$, $\forall i = 1, \cdots, m$, 则对于 $x_1, \cdots, x_m$ 的任意凸组合 $x = \sum\limits_{i=1}^{m} \lambda_i x_i$, 有

$$\langle y, x\rangle = \left\langle y, \sum_{i=1}^{m} \lambda_i x_i \right\rangle = \sum_{i=1}^{m} \lambda_i \langle y, x_i\rangle \le 1.$$

从而 $\langle y, x\rangle \le 1$, $\forall x \in P$, 即 $y \in P^o$. ∎

假定 S 是 $\mathbb{R}^n$ 中的一个多面凸集：

$$S = \bigcap_{i=1}^{m} \{x \in \mathbb{R}^n \mid \langle x, y_i\rangle \le 1\}.$$

记 $P = \operatorname{conv}\{y_1, \cdots, y_m\}$, 则由上述定理可知 $P^o = S$. 此外，若还有 $0 \in P$, 则依定理 3.3.4, $S^o = P$.

由任意有限集 $\{x_1, \cdots, x_m\}$ 的凸包 $\operatorname{conv}\{x_1, \cdots, x_n\}$ 构成的多面体未必是吸收的.

定理 3.3.10 为了 $\mathbb{R}^n$ 中的凸多面体 $P = \operatorname{conv}\{x_1, \cdots, x_m\}$ 是吸收的，必须且只须对于任意非零向量 $x \in \mathbb{R}^n$, 诸 $\langle x, x_i\rangle$ 中必定有严格正的也必定有严格负的；同样地，若多面体 P 表示成

$$S = \bigcap_{i=1}^{m} \{x \in \mathbb{R}^n \mid \langle x, y_i\rangle \le \alpha_i\},$$

则为了 S 是吸收的，必须且只须每一个半空间 $\{x \in \mathbb{R}^n \mid \langle x, y_i\rangle \le \alpha_i\}$ 可以写成 $\{x \in \mathbb{R}^n \mid \langle x, z_i\rangle \le 1\}$, 并且 $\operatorname{conv}\{z_1, \cdots, z_m\}$ 是吸收的多面体.

证明 我们已经知道，P 为吸收集的充分必要条件是 $0 \in \operatorname{int} P$. 但 $\{x \in \mathbb{R}^n \mid \langle x, x_i\rangle = 0\}$ 正好是通过原点的超平面，显然 P 不可能位于这个超平面的同一侧，因此定理的第一个结论成立.

注意，为了 $0 \in \operatorname{int} S$ (即 S 是吸收的), 必须且只须每一个半空间 $\{x \in \mathbb{R}^n \mid \langle x, x_i\rangle \le \alpha_i\}$ 包含 0 为其内点. 因

此必须且只须 $\alpha_i > 0$, 从而 $\{x \in \mathbb{R}^n \mid \langle x, x_i \rangle \le \alpha_i\}$ 可以写成 $\{x \in \mathbb{R}^n \mid \langle x, z_i \rangle \le 1\}$. 由此从推论 3.3.5 得知定理的第二个结论成立. ∎

最后我们给出几个对偶锥的例子.

例 3.3.4 设 K 是 $\mathbb{R}^n$ 的线性子空间, 则 K 的对偶锥 K^* 等于 K 的正交补, 即 $K^* = K^{\perp}$. 这是极化锥的相应的结论.

例 3.3.5 设 $y \in \mathbb{R}^n$, 并令

$$K_1 = \{x \in \mathbb{R}^n \mid \langle x, y \rangle = 0\},$$

$$K_2 = \{x \in \mathbb{R}^n \mid \langle x, y \rangle \ge 0\},$$

$$K_3 = \{x \in \mathbb{R}^n \mid \langle x, y \rangle > 0\}.$$

那么 K_1 为子空间, K_2 和 K_3 为凸锥, 它们的对偶锥分别为

$$K_1^* = \{\alpha y \mid \lambda \in \mathbb{R}\},$$
$$K_2^* = \{\lambda y \mid 0 \le \lambda < \infty\},$$
$$K_3^* = \begin{cases} \mathbb{R}^n, & \text{若 } y = 0, \\ K_2^*, & \text{若 } y \ne 0. \end{cases}$$

事实上, 依定理 3.3.11,

$$K_1^* = \{z \in \mathbb{R}^n \mid \langle x, z \rangle = 0,\ \forall \langle x, y \rangle = 0\}.$$

如果 $y = 0$, 则 $K_1 = \mathbb{R}^n$, 从而 $K_1^* = \{0\}$; 如果 $y \ne 0$, 可取 $x_0 \in \mathbb{R}^n$ 使得 $\langle x_0, y \rangle \ne 0$. 令 $\lambda = \langle x_0, z \rangle / \langle x_0, y \rangle$, 则显然

$$\langle x - (\langle x, y \rangle / \langle x_0, y \rangle) x_0, y \rangle = 0, \quad \forall x \in \mathbb{R}^n.$$

于是 $z \in K_1^*$ 等价于

$$\langle x - (\langle x, y \rangle / \langle x_0, y \rangle) x_0, z \rangle = 0, \quad \forall x \in \mathbb{R}^n,$$

即 $z = \lambda y$. 这样我们证明了 K_1^* 的表达式.

由于 $K_1 \subset K_2$, 故 $K_2^* \subset K_1^*$. 若 $z = \lambda y \in K_2^*$, 则 $\langle z, x\rangle = \lambda\langle y, x\rangle \geq 0$, $\forall x \in K_2$. 但为此必须且只须 $\lambda \geq 0$.

最后, 若 $y = 0$, 则 $K_3 = \emptyset$, $K_3^* = \mathbb{R}^n$; 而若 $y \neq 0$, 则显然 $\operatorname{cl} K_3 = K_2$. 因此 $K_3^* = K_2^*$.

习　题　3.3

3.3.1 设 K 为 $\mathbb{R}^n$ 的非负象限:

$$K = \{(\xi_1, \cdots, \xi_n)^\tau \mid \xi_i \geq 0,\ 1 \leq i \leq n\},$$

试证: $K^o = -K$ (非正象限).

3.3.2 设 $\{a_i \mid i \in I\} \subset \mathbb{R}^n$ 为一族向量, 并设 K 是由向量族 $\{a_i \mid i \in I\}$ 生成的凸锥, 试证:

$$K^o = \{x^* \in \mathbb{R}^n \mid \langle a_i, x^*\rangle \leq 0,\ i \in I\}.$$

3.3.3 设

$$\begin{aligned}
M_1 &= \{(\xi_1, \cdots, \xi_n)^\tau \in \mathbb{R}^n \mid \xi_i \geq 0,\ \xi_1 + \cdots + \xi_n \leq 1\},\\
M_2 &= \{(\xi_1, \cdots, \xi_n)^\tau \in \mathbb{R}^n \mid |\xi_1| + \cdots + |\xi_n| \leq 1\},\\
M_3 &= \{(\xi_1, \xi_2)^\tau \in \mathbb{R}^2 \mid (\xi_1 - 1)^2 + \xi_2^2 \leq 1\},\\
M_4 &= \{(\xi_1, \xi_2)^\tau \in \mathbb{R}^2 \mid \xi_1 \leq 1 - (1 + \xi_2^2)^{1/2}.
\end{aligned}$$

求证:

$$\begin{aligned}
M_1^o &= \{(\eta_1, \cdots, \eta_n)^\tau \in \mathbb{R}^n \mid \eta_i \leq 1,\ i = 1, \cdots, n\},\\
M_2^o &= \{(\eta_1, \cdots, \eta_n)^\tau \in \mathbb{R}^n \mid |\eta_i| \leq 1,\ i = 1, \cdots, n\},\\
M_3^o &= \{(\eta_1, \eta_2)^\tau \in \mathbb{R}^2 \mid \eta_1 \leq (1 - \eta_2^2)/2\},\\
M_4^o &= \operatorname{conv}(P \cup \{0\}),\quad P = \{(\eta_1, \eta_2)^\tau \mid \eta_1 \geq (1 + \eta_2^2)/2\}.
\end{aligned}$$

3.3.4 设 $M \subset \mathbb{R}^n$ 为非空集, 试证:

(1)$M^{ooo} = M^o$;

(2) M 有界 $\Longleftrightarrow 0 \in \operatorname{int} M^o$;

(3) M^o 有界 $\iff 0 \in \operatorname{int}(\operatorname{conv} M)$.

3.3.5 设 $\{M_i \mid i \in I\}$ 为 $\mathbb{R}^n$ 的一族非空紧凸集，假定 $0 \in \operatorname{int} M_i$, $\forall i \in I$. 试证：

$$\left(\bigcap_{i\in I} M_i\right)^o = \operatorname{cl}\left(\operatorname{conv}\left(\bigcup_{i\in I} M_i^o\right)\right).$$

3.3.6 设 $M \subset \mathbb{R}^n$ 为一空凸子集，假定

$$\langle x, z\rangle \le \sup\{\langle y, z\rangle \mid y \in M\}, \quad \forall z \in \mathbb{R}^n,$$

求证：$x \in M$.

3.3.7 设 A 为 $\mathbb{R}^n$ 中一正定矩阵，S 为由 A 决定的椭球体：$S = \{z \in \mathbb{R}^n \mid \langle z, Az\rangle \le 1\}$. 试证：

$$\gamma_S(x) = \|A^{1/2}x\|, \quad \forall x \in \mathbb{R}^n,$$

$$\gamma_{S^o}(x) = \|A^{-1/2}x\|, \quad \forall x \in \mathbb{R}^n.$$

$$\inf_{\alpha\in\mathbb{R}} \gamma_{S^o}(x+\alpha y) = \frac{\left(\langle x, A^{-1}x\rangle\langle y, A^{-1}y\rangle - (\langle y, A^{-1}y\rangle)^2\right)^{1/2}}{(\langle y, A^{-1}y\rangle)^{1/2}}.$$

3.3.8 设 $K \subset \mathbb{R}^n$ 为紧凸集，假定 $0 \in \operatorname{int} K$, F 为 K 的一个面. 定义 K^o 的子集 $\hat{F}$:

$$\hat{F} = \{y \in K^o \mid \langle y, x\rangle = 1, \ \forall x \in F\}.$$

试证：$\hat{F}$ 是 K^o 的一个面.

3.4 对偶运算

假定我们对给定的一些凸函数 $f_1, \cdots, f_m$ 作某种运算，如加法，数乘等等，所得到的函数的共轭与 $f_1^*, \cdots, f_m^*$ 有什么关系呢？实际上，在许多情况下，这种共轭对应会把一种运算转换成另一种运算.

我们从一些简单的情形着手讨论. 设 g 为 $\mathbb{R}^n$ 上任一凸函数，取 $a \in \mathbb{R}^n$，作 g 的平移：$f(x) = g(x-a)$，则我们知道

$$f^*(x^*) = g^*(x^*) + \langle a, x^* \rangle, \quad \forall x^* \in I\!R^n.$$

f^* 作为 g^* 与线性函数之和的共轭是 f 的二次共轭 f^{**}, 容易算出

$$f^{**}(x) = g^{**}(x - a), \quad \forall x \in I\!R^n.$$

特别是，若 g 还是闭的，则 $g(x-a)$ 与 $g^*(x^*) + \langle a, x^* \rangle$ 互为共轭.

再如，对于实数 α 和凸函数 g, 我们有

$$(g + \alpha)^* = g^* - \alpha.$$

对于 $a \in I\!R^n$ 和凸集 $M \subset I\!R^n$, 我们知道，平移集合 $M + a$ 的承托函数是

$$\sigma(x^*|M + a) = \sigma(x^*|M) + \langle a, x^* \rangle.$$

实际上，这一事实是上面指出的规则的特殊情形：示性函数 $g = \delta(\cdot|M)$ 的共轭是 M 的承托函数；而 M 的平移对应于其示性函数的平移.

凸函数的左右非负数乘法也是互为共轭的.

定理 3.4.1 设 f 为 $I\!R^n$ 上的真凸函数. 那么

$$(\lambda f)^* = f^*\lambda, \ (f\lambda)^* = \lambda f^*, \quad \forall 0 \le \lambda < \infty.$$

证明 当 $\lambda > 0$ 时直接从共轭函数定义推出. 当 $\alpha = 0$ 时它反映了如下一个简单的事实；常值函数 0 是示性函数 $\delta(\cdot|0)$ 的共轭函数. ∎

推论 3.4.2 设 $M \subset I\!R^n$ 为非空凸集，那么

$$\sigma(x^*|\lambda M) = \lambda\sigma(x^*|M), \quad \forall 0 \le \lambda < \infty.$$

证明 取 $f(x) = \delta(x|M)$. ∎

在进一步讨论之前我们先证明一个引理.

引理 3.4.3 设 $L \subset \mathbb{R}^n$ 为一子空间, f 为 $\mathbb{R}^n$ 上一真凸函数. 那么, $L \cap \mathrm{ri}\,(\mathrm{dom}\,f) \neq \emptyset$, 当且仅当不存在任何向量 $x^* \in L^\perp$ 使得 $(f^*0^+)(x^*) \leq 0$ 和 $(f^*0^+)(-x^*) > 0$.

证明 由于 L 是相对开集, 故 $L \cap \mathrm{ri}\,(\mathrm{dom}\,f) = \emptyset$ 等价于存在一超平面把 L 和 $\mathrm{dom}\,f$ 真分离, 即存在一向量 $x^* \in \mathbb{R}^n$ 使得

$$\inf\{\langle x, x^*\rangle \mid x \in L\} \geq \sup\{\langle x, x^*\rangle \mid x \in \mathrm{dom}\,f\},$$

$$\sup\{\langle x, x^*\rangle \mid x \in L\} > \inf\{\langle x, x^*\rangle \mid x \in \mathrm{dom}\,f\}.$$

(见引理 1.5.2) 上述 $\mathrm{dom}\,f$ 上的上、下确界分别是

$$(f^*0^+)(x^*) \text{ 和 } -(f^*0^+)(-x^*),$$

这是因为, 根据定理 3.2.11, f^*0^+ 是 $\mathrm{dom}\,f$ 的承托函数. 另一方面, L 上 $\langle x, x^*\rangle$ 的下确界当 $x^* \in L^\perp$ 时为 0, 而当 $x^* \notin L^\perp$ 时为 $-\infty$. 因此上述两个关于 x^* 的极值条件等价于 $x^* \in L^\perp$, $0 \geq (f^*0^+)(x^*)$ 和 $0 > -(f^*0^+)(-x^*)$. ∎

推论 3.4.4 设 $A: \mathbb{R}^n \to \mathbb{R}^m$ 为线性变换. 设 g 为 $\mathbb{R}^m$ 上的真凸函数. 那么为了不存在向量 $y^* \in \mathbb{R}^m$ 使得

$$A^*y^* = 0,\ (g^*0^+)(y^*) \leq 0 \text{ 和 } (g^*0^+)(-y^*) > 0$$

同时满足, 必须且只须至少有一个向量 $x \in \mathbb{R}^n$ 使得 $Ax \in \mathrm{ri}\,(\mathrm{dom}\,g)$.

证明 令 $L = \{Ax \mid x \in \mathbb{R}^n\}$, 即 L 为 A 的值域子空间, 则

$$L^\perp = \{y^* \in \mathbb{R}^m \mid A^*y^* = 0\}.$$

然后应用引理 3.4.3 于 L 和 g, 即得所要结论. ∎

推论 3.4.5 设 $f_1, \cdots, f_m$ 是 $\mathbb{R}^n$ 上的 m 个真凸函数. 那么为了不存在向量 $x_1^*, \cdots, x_m^* \in \mathbb{R}^n$, 使得

$$x_1^* + \cdots + x_m^* = 0,$$

$$(f_1^* 0^+)(x_1^*) + \cdots + (f_m^* 0^+)(x_m^*) \le 0,$$

$$(f_1^* 0^+)(-x_1^*) + \cdots + (f_m^* 0^+)(-x_m^*) > 0,$$

必须且只须

$$\mathrm{ri}\,(\mathrm{dom}\, f_1) \cap \cdots \cap \mathrm{ri}\,(\mathrm{dom}\, f_m) \ne \emptyset.$$

证明 把 $\mathbb{R}^{mn}$ 看成乘积空间 $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$, 对于 $\mathbb{R}^{mn}$ 中的元 $x = (x_1, \cdots, x_m)$, $x^* = (x_1^*, \cdots, x_m^*)$, $x_i, x_i^* \in \mathbb{R}^n$, $1 \le i \le m$, 规定内积为

$$\langle x, x^* \rangle = \langle x_1, x_1^* \rangle + \cdots + \langle x_m, x_m^* \rangle.$$

今定义 $\mathbb{R}^{mn}$ 上的函数 f:

$$f(x_1, \cdots, x_m) = \sum_{k=1}^{m} f_k(x_k), \quad \forall\, x_i \in \mathbb{R}^n,\ 1 \le i \le m.$$

f 显然是凸函数, 并且由于假设

$$\mathrm{ri}\,(\mathrm{dom}\, f_1) \cap \cdots \cap \mathrm{ri}\,(\mathrm{dom}\, f_m) \ne \emptyset,$$

f 还是真凸的. 容易看出, 对于 $x_i \in \mathbb{R}^n, 1 \le i \le m$,

$$f^*(x_1^*, \cdots, x_m^*) = f_1^*(x_1^*) + \cdots + f_m^*(x_m^*),$$

并且 f^* 的回收函数 $f^* 0^+$ 是

$$(f^* 0^+)(x_1^*, \cdots, x_m^*) = (f_1^* 0^+)(x_1^*) + \cdots + (f_m^* 0^+)(x_m^*).$$

定义子空间

$$L = \{(x_1, \cdots, x_m) \in \mathbb{R}^{mn} \mid x_1 = x_2 = \cdots = x_m \in \mathbb{R}^n\},$$

则其正交补是

$$L^{\perp} = \{x^* = (x_1^*, \cdots, x_m^*) \mid x_1^* + \cdots + x_m^* = 0\}.$$

然后应用引理 3.4.3 于 f 和 L 的即得所要结论.

定理 3.4.6 设 $A : \mathbb{R}^n \to \mathbb{R}^m$ 为线性变换，那么对于 $\mathbb{R}^n$ 上的任意凸函数 f, 有

$$(Af)^* = f^* A^*,$$

而对于 $\mathbb{R}^m$ 上的任意凸函数 g, 则有

$$((\mathrm{cl}\, g)A)^* = \mathrm{cl}\,(A^* g^*).$$

此外，如果存在 $x \in \mathbb{R}^n$ 使得 $Ax \in \mathrm{ri}\,(\mathrm{dom}\, g)$, 则上述第二式中的闭包运算可以略去，并且

$$(gA)^*(x^*) = \inf\{g^*(y^*) \mid A^* y^* = x^*\},$$

其中对于每一 $x^* \in \mathbb{R}^m$, 下确界被达到 (空集上的下确界约定为 $+\infty$).

证明 直接计算可以证明第一式:

$$\begin{aligned}
(Af)^*(y^*) &= \sup\{\langle y, y^* \rangle - \inf_{Ax=y} f(x) \mid y \in \mathbb{R}^m\} \\
&= \sup\{\sup_{Ax=y} \{\langle y, y^* \rangle - f(x) \mid y \in \mathbb{R}^m\} \\
&= \sup\{\langle Ax, y^* \rangle - f(x) \mid x \in \mathbb{R}^n\} \\
&= \sup\{\langle x, A^* y^* \rangle - f(x) \mid x \in \mathbb{R}^n\} \\
&= f^*(A^* y^*).
\end{aligned}$$

应用这一关系于 A^* 和 g^* 得到

$$(A^*g^*)^* = g^{**}A^{**} = (\mathrm{cl}\, g)A,$$

因此

$$((\mathrm{cl}\, g)A)^* = (A^*g^*)^{**} = \mathrm{cl}\,(A^*g^*).$$

现在假定存在 $x \in I\!R^n$, 使得 $Ax \in \mathrm{ri}\,(\mathrm{dom}\, g)$. 如果对某个 $y \in I\!R^m$, $g(y) = -\infty$, 则依据定理 2.2.11, g 在整个 $\mathrm{ri}\,(\mathrm{dom}\, g)$ 上恒等于 $-\infty$, 从而 g^* 和 $(gA)^*$ 恒等于 $+\infty$. 因此，我们假定 $g(y) > -\infty$, $\forall y \in I\!R^m$. 于是根据定理 2.6.8, $(\mathrm{cl}\, g)A = \mathrm{cl}\,(gA)$. 两边取共轭得到 $((\mathrm{cl}\, g)A)^* = (gA)^*$. 另一方面，由于 $Ax \in \mathrm{ri}\,(\mathrm{dom}\, g)$, 因此推论 3.4.4 断定 g^* 和 A^* 满足定理 2.6.5 的条件，从而 $\mathrm{cl}\,(A^*g^*) = A^*g^*$, 并且 A^*g^* 定义中的下确界被达到. ∎

注意，上述定理的最后一个结论是说：

$$\sup\{\langle x, x^*\rangle - g(Ax) \mid x \in I\!R^n\} = \inf\{g^*(y^*) \mid A^*y^* = x^*\}.$$

类似于这样的 inf = sup 的公式在极值问题中占有重要的位置.

我们继续讨论加法和下端卷积的对偶关系.

定理 3.4.7 设 $f_1, \cdots, f_m$ 为 $I\!R^n$ 上的真凸函数. 那么

$$(f_1 \square \cdots \square f_m)^* = f_1^* + \cdots + f_m^*,$$

$$(\mathrm{cl}\, f_1 + \cdots + \mathrm{cl}\, f_m)^* = \mathrm{cl}\,(f_1^* \square \cdots \square f_m^*).$$

此外，如果 $\bigcap\limits_{r=1}^{m} \mathrm{ri}\,(\mathrm{dom}\, f_i) \neq \emptyset$, 则第二式中的闭包运算可以去掉，并且

$$\begin{aligned}&(f_1 + \cdots + f_m)^*(x^*)\\ &= \inf\{f_1(x_1^*) + \cdots + f_m^*(x_m^*) \mid x_1^* + \cdots + x_m^* = x^*\},\end{aligned}$$

这是对每一 x^*, 下确界都能达到.

证明　依照定义,

$$
\begin{aligned}
&(f_1\Box\cdots\Box f_m)^*(x^*)\\
=&\sup_{x\in \mathbb{R}^n}\left\{\langle x,x^*\rangle-\inf_{x_1+\cdots+x_m=x}\{f_1(x_1)+\cdots+f_m(x_m)\}\right\}\\
=&\sup_{x\in \mathbb{R}^n}\sup_{x_1+\cdots+x_m=x}\{\langle x,x^*\rangle-f_1(x_1)-\cdots-f_m(x_m)\}\\
=&\sup_{x_1,\cdots,x_m\in \mathbb{R}^n}\{\langle x_1,x^*\rangle+\cdots+\langle x_m,x^*\rangle\\
&-f_1(x_1)-\cdots-f_m(x_m)\}\\
=&f_1^*(x^*)+\cdots+f_m^*(x_m^*).
\end{aligned}
$$

据此又可推出

$$(f_1^*\Box\cdots\Box f_m^*)^*=f_1^{**}+\cdots+f_m^{**}=\mathrm{cl}\,f_1+\cdots+\mathrm{cl}\,f_m,$$

因此

$$(\mathrm{cl}\,f_1+\cdots+\mathrm{cl}\,f_m)^*=(f_1^*\Box\cdots\Box f_m^*)^{**}=\mathrm{cl}\,(f_1^*\Box\cdots\Box f_m^*).$$

现在如果 $\bigcap_{i=1}^m \mathrm{ri}\,(\mathrm{dom}\,f_i)\neq\emptyset$, 则根据定理 2.6.7,

$$\mathrm{cl}\,f_1+\cdots+\mathrm{cl}\,f_m=\mathrm{cl}\,(f_1+\cdots+f_m).$$

上式右端的共轭是 $(f_1+\cdots+f_m)^*$. 同时在 $\bigcap_{i=1}^m \mathrm{ri}\,(\mathrm{dom}\,f_i)\neq\emptyset$ 的假设下，推论 2.6.6 条件满足，从而下端卷积 $f_1^*\Box\cdots\Box f_m^*$ 是闭的，并且其定义中的下确界被达到. ∎

现在我们应用定理 3.4.7 来计算距离函数的共轭. 设 M 为 $\mathbb{R}^n$ 的一非空凸集，令

$$f(x)=d(x|M)=\inf\{\|x-y\|\mid y\in M\},\quad \forall x\in \mathbb{R}^n.$$

如果令 $f_1(x)=\|x\|$, $f_2(x)=\delta(x|M)$, 则 $f=f_1\Box f_2$. 于是

$$f^*(x^*)=f_1^*(x^*)+f_2^*(x^*)=\begin{cases}\sigma(x^*|M), & \forall\|x\|\le 1,\\ +\infty, & \text{其余处}.\end{cases}$$

下一个例子会在逼近论中遇到. 考察函数

$$f(x)=\inf\{\|x-\zeta_1y_1-\cdots-\zeta_my_m\|_\infty \mid \zeta_i\in I\!R\},$$

其中 $y_1,\cdots,y_m$ 为 $I\!R^n$ 中的给定元, 而

$$\|z\|_\infty=\max\{|\zeta_j| \mid 1\le j\le n\},\quad \forall\, z=(\zeta_1,\cdots,\zeta_n)^\tau\in I\!R^n$$

于是若令

$$f_1(x)=\|x\|_\infty,\quad f_2(x)=\delta(x|L),$$

其中 $L=\operatorname{span}\{y_1,\cdots,y_m\}$, 则 $f=f_1\Box f_2$. 注意 f_1 是集合

$$M=\{x=(\xi_1,\cdots,\xi_n)^\tau \mid |\xi_1|+\cdots+|\xi_m|\le 1\}$$

的承托函数, 故 f_1^* 是 M 的示性函数. 另一方面, f_2^* 是正交补空间

$$L^\perp=\{x^* \mid \langle x^*,y_i\rangle=0,\ i=1,\cdots,m\}$$

的示性函数. 因此

$$f^*=f_1^*+f_2^*=\delta(\cdot|M\cap L^\perp).$$

定理 3.4.8 设 $\{f_i \mid i\in I\}$ 是 $I\!R^n$ 上的一族真凸函数, I 为任一指标集. 那么

$$\Big(\operatorname{conv}\{f_i \mid i\in I\}\Big)^*=\sup\{f_i^* \mid i\in I\},$$

$$\Big(\sup\{\operatorname{cl} f_i \mid i\in I\}\Big)^*=\operatorname{cl}\Big(\operatorname{conv}\{f_i^* \mid i\in I\}\Big).$$

如果 I 是一个有穷指标集, 并且 $\operatorname{cl}(\operatorname{dom} f_i)$ 为同一集合 M, 则上述公式中的闭包运算可以略去. 此外, 在这种情形下,

$$\Big(\sup\{f_i \mid i\in I\}\Big)^*(x^*)$$

$$=\inf\left\{\sum_{i\in I}\lambda_i f_i^*(x_i^*)\;\middle|\;\sum_{i\in I}\lambda_i x_i^*=x^*\right\},\qquad(3.4.1)$$

这里对每一 $x^* \in \mathbb{R}^n$, 式 (3.4.1) 中的下确界取在一切凸组合 $\sum_{i\in I} \lambda_i x_i^* = x^*$ 上，并且下确界被达到.

证明 记 $f = \operatorname{conv}\{f_i | i \in I\}$, 我们知道，$(x^*, \mu^*) \in \operatorname{epi} f^*$ 等价于相应的仿射函数 $h = \langle \cdot, x^* \rangle - \mu^*$ 满足 $h \le f$. 但 $h \le f$ 又等价于 $h \le f_i$, $\forall i \in I$, 于是 $\mu^* \ge f^*(x^*)$, 当且仅当 $\mu^* \ge f_i^*(x^*)$, $\forall i \in I$. 这正是所要证明的第一个式子. 将此公式应用于 f_i^*, 我们得到

$$\Big(\operatorname{conv}\{f_i^* \mid i \in I\}\Big)^* = \sup\{f_i^{**} \mid i \in I\} = \sup\{\operatorname{cl} f_i \mid i \in I\}.$$

然后再对上式两端取共轭，得到

$$\begin{aligned}\Big(\sup\{\operatorname{cl} f_i \mid i \in I\}\Big)^* &= \Big(\operatorname{conv}\{f_i^* \mid i \in I\}\Big)^{**} \\ &= \operatorname{cl}\Big(\operatorname{conv}\{f_i^* \mid i \in I\}\Big).\end{aligned}$$

如果存在一点 $z \in \bigcap_{i\in I} \operatorname{ri}(\operatorname{dom} f_i)$ 使得 $\sup_{i\in I} f_i(z) < \infty$, 则依据定理 2.6.9,

$$\begin{aligned}\Big(\sup\{\operatorname{cl} f_i \mid i \in I\}\Big)^* &= \Big(\operatorname{cl}(\sup\{f_i \mid i \in I\})\Big)^* \\ &= \Big(\sup\{f_i \mid i \in I\}\Big)^*.\end{aligned}$$

特别是当 I 为有穷集，并且 $\operatorname{cl}(\operatorname{dom} f_i) = M$, $\forall i \in I$, 就是这种情形. 这时集合 $\operatorname{dom} f_i$ 的承托函数是回收函数 $f_i^* 0^+$ (见定理 3.2.11), 并且都等于 $\sigma(\cdot|M)$, 因此依推论 2.6.13, $\operatorname{conv}\{f_i^* \mid i \in I\}$ 是闭的，并且由 (3.4.1) 中的下确界式子给出. ▮

推论 3.4.9 设 $\{M_i \mid i \in I\}$ 是 $\mathbb{R}^n$ 中一族非空凸集，并记 $D = \operatorname{conv}\{M_1 \cup \cdots \cup M_m\}$, 那么

$$\sigma(\cdot|D) = \sup\{\sigma(\cdot|M_i) \mid i \in I\}.$$

此外，若 $M = \bigcap\limits_{i\in I} \operatorname{cl} M_i$，则

$$\sigma(\cdot|M) = \operatorname{cl}\left(\operatorname{conv}\{\sigma(\cdot|M_i) \mid i \in I\}\right).$$

证明 根据定理 3.4.8, 取 $f_i = \delta(\cdot|M_i)$. ▌

推论 3.4.10 设 $\{M_i \mid i \in I\}$ 是 $\mathbb{R}^n$ 中一族凸集, 那么

$$\left(\operatorname{conv}\{M_i \mid i \in I\}\right)^o = \cap\{M_i^o \mid i \in I\},$$

$$\left(\cap\{\operatorname{cl} M_i \mid i \in I\}\right)^o = \operatorname{cl}\left(\operatorname{conv}\{M_i^o \mid i \in I\}\right).$$ ▌

在结束本节之前，我们把有关凸函数和凸集运算的几个重要的对偶关系简单归纳于下：

(1) 共轭运算 $f \to f^*$ 在 $\mathbb{R}^n$ 上的闭真凸函数类中是一对一的.

(2) 对于 $\mathbb{R}^n$ 上的真凸函数 f, $f^{**} = \operatorname{cl} f$.

(3) 对于 $\mathbb{R}^n$ 上的真凸函数 f,

$$f(x) + f^*(x^*) \geq \langle x, x^*\rangle, \quad \forall\, x, x^* \in \mathbb{R}^n.$$

(4) 对于 $\mathbb{R}^n$ 上的真凸函数 $f_1, f_2, \cdots, f_m$,

$$(f_1 \square \cdots \square f_m)^* = f_1^* + \cdots + f_m^*,$$

$$(\operatorname{cl} f_1^* + \cdots + \operatorname{cl} f_m^*)^* = \operatorname{cl}(f_1^* \square \cdots \square f_m^*).$$

(5) 对于 $\mathbb{R}^n$ 上的真凸函数族 $\{f_i \mid i \in I\}$,

$$\left(\operatorname{conv}\{f_i \mid i \in I\}\right)^* = \sup\{f_i^* \mid i \in I\},$$

$$\left(\{\operatorname{cl} f_i \mid i \in I\}\right)^* = \operatorname{cl}\left(\{f_i^* \mid i \in I\}\right).$$

(6) 对于 $\mathbb{R}^n \to \mathbb{R}^m$ 的线性变换 A, $\mathbb{R}^n$ 上的凸函数 f 和 $\mathbb{R}^m$ 上的凸函数 g,

$$(Af)^* = f^* A^*, \quad \left((\operatorname{cl} g)A\right)^* = \operatorname{cl}(A^* g^*).$$

(7) 对于 $I\!R^n$ 中的一族凸集 $\{M_i \mid i \in I\}$,

$$\left(\operatorname{conv}\{M_i \mid i \in I\}\right)^o = \cap\{M_i^o \mid i \in I\},$$

$$\left(\cap\{\operatorname{cl} M_i \mid i \in I\}\right)^o = \operatorname{cl}\left(\operatorname{conv}\{M_i^o \mid i \in I\}\right).$$

习　题　3.4

3.4.1 设 $A: I\!R^n \to I\!R^m$ 为线性变换. 试证: 对于 $I\!R^n$ 中的任意凸集 M, 式

$$\sigma(y|AM) = \sigma(A^*y|M), \quad \forall y \in I\!R^m,$$

成立, 而对于任意凸集 $N \subset I\!R^m$, 式

$$\sigma(x|A^{-1}(\operatorname{cl} N)) = \operatorname{cl}(A^*\sigma(x|N)), \quad \forall x \in I\!R^n$$

成立, 此外, 如果存在一个 $x \in I\!R^n$ 使得 $Ax \in \operatorname{ri} N$, 则上式中闭包运算可以省略, 并且

$$\sigma(x|A^{-1}N) = \inf\{\sigma(y|N) \bigm| A^*y = x\},$$

这里对于每一 $x \in I\!R^n$, 下确界被达到. (约定在空集 $\emptyset$ 上的下确界为 $+\infty$)

3.4.2 设 $A: I\!R^n \to I\!R^m$ 为线性变换. 试证: 对于任意凸集 $M \subset I\!R^n$, 式

$$(AM)^o = A^{*-1}(M^o)$$

成立, 而对于任意凸集 $N \subset I\!R^m$, 式

$$(A^{-1}(\operatorname{cl} N))^o = \operatorname{cl}(A^*(N^o))$$

成立. 此外, 若存在 $x \in I\!R^n$ 使得 $Ax \in \operatorname{ri} N$, 则上式中的闭包运算可以省略. 这里 K^o 表示 K 的极化集.

3.4.3 设 $M_1, \cdots, M_m$ 是 $I\!R^n$ 中的非空凸集. 试证:

$$\sigma(\cdot|M_1 + \cdots + M_m) = \sigma(\cdot|M_1) + \cdots + \sigma(\cdot|M_m),$$

$$\sigma(\cdot|\operatorname{cl} M_1 \cap \cdots \cap \operatorname{cl} M_m) = \operatorname{cl}(\sigma(\cdot|M_1)\square \cdots \square\sigma(\cdot|M_m)).$$

此外, 如果诸集合 $\operatorname{ri} M_1, \cdots, \operatorname{ri} M_m$ 相交, 则上式中的闭包运算可以省略, 并且有

$$\sigma(x|M_1\cap\cdots\cap M_m)=\inf\left\{\sum_{k=1}^m\sigma(x_k|M_k)\,\middle|\,\sum_{k=1}^m x_k=x\right\},$$

这里的下确界对每个 $x\in\mathbb{R}^n$ 都被达到.

3.4.4 设 $K_1,\cdots,K_m$ 是 $\mathbb{R}^n$ 中的非空凸锥. 试证:

$$(K_1+\cdots+K_m)^o=K_1^o\cap\cdots\cap K_m^o,$$
$$(\operatorname{cl}K_1\cap\cdots\cap\operatorname{cl}K_m)^o=\operatorname{cl}(K_1^o+\cdots+K_m^o).$$

此外, 如果诸锥 $\operatorname{ri}K_1,\cdots,\operatorname{ri}K_m$ 相交, 则上式中的闭包运算可以省略.

3.5 无界多面凸集*

本节讨论无界多面凸集的基本性质. 在此之前, 我们先讨论一下凸集的表现问题. 我们在 1.6 节中曾指出 $\mathbb{R}^n$ 中的每个闭凸集 M 都是 $\mathbb{R}^n$ 中包含 M 的所有闭半空间之交. 实际上, 这样的闭半空间的数目可以大大减少. 我们称超平面 $H\subset\mathbb{R}^n$ 与闭凸集 $M\subset\mathbb{R}^n$ 在一点 $x\in M$ 相切, 是指 H 是 M 在 x 处的唯一的承托超平面. M 的切半空间是指包含 M 的一闭半空间, 并且其边界超平面 (实际上半空间的边界本身是一个超平面) 与 M 在某点相切. 这里的切超平面实际上是 1.9 节中定义的暴露点的一种对偶概念.

定理 3.5.1 设 M 是 $\mathbb{R}^n$ 中的一 n 维闭凸集. 那么 M 是其所有切半空间之交.

证明 设 G 表示 M 的承托函数 $\sigma(\cdot|M)$ 的上图, 显然 G 是 $\mathbb{R}^{n+1}$ 中不止含原点的闭凸锥. 由于 $\dim M=n$, M 的内部非空, 故

$$-\sigma(-x^*|M)<\sigma(x^*|M),\quad \forall\, x^*\in\mathbb{R}^n,\, x^*\neq 0.$$

因此 G 不可能包含通过原点的直线. 依据推论 1.9.11, 有

$G = \mathrm{cl}(\mathrm{conv}\, S)$, 其中 S 为 G 的所有暴露射线的并. 注意, 根据定理 3.2.4,

$$x \in M \Longleftrightarrow \langle x, x^* \rangle \le \sigma(x^*|M), \quad \forall x^* \in \mathbb{R}^n.$$

因此, 对于 $x \in M$, 相应的线性泛函 $\langle x, \cdot \rangle$ 的上图含有 G 的 "非垂直" 的暴露射线; 反之也一样. 另一方面, M 又是这样一些闭半空间 $\{x \in \mathbb{R}^n \mid \langle x, x^* \rangle \le \alpha\}$ 之交: 这里的 (x^*, α) 使得 $\{\lambda(x^*, \alpha) \mid \lambda \ge 0\}$ 是 G 的 "非垂直" 暴露射线. 后一个条件意味着存在 G 的某个 "非垂直" 的承托超平面 (即某个线性函数 $\langle \cdot, y \rangle$ 的图像), 它与 G 之交就是由 (x^*, α) 产生的射线. 换句话说, 存在某个 $y \in M$, 使得 $\langle y, x^* \rangle = \sigma(x^*|M) = \alpha$, 但当 y^* 不等于 x^* 的非负倍数时, 有 $\langle y, y^* \rangle < \sigma(y^*|M)$. 这正是说半空间 $\{\mathbb{R}^n \mid \langle x, x^* \rangle \le \alpha\}$ 与 M 相切于点 y. 因此 M 是所有这样的切半空间的交. ▮

我们称一个凸集是有限生成的, 是指它是由有穷多个点和方向构成的某个集合的凸包. 下一个定理是说, 多面凸集就是有限生成的凸集. 这一事实在几何上是相当直观的, 但其证明却并不是初等的. 这里我们基于 1.9 节中建立的关于凸集面的结果来证明这一定理.

定理 3.5.2 设 M 是 $\mathbb{R}^n$ 中的凸集. 那么如下三个命题是等价的:

(1) M 是多面凸集;

(2) M 是闭的, 并且只有有穷多个面;

(3) M 是有限生成的.

证明 (1)$\Longrightarrow$(2): 设 M 是 $\mathbb{R}^n$ 中 m 个闭半空间 $G_1, \cdots, G_m$ 之交, 并设 F 为 M 的一个非空面. 设 H_k 为由 G_k 的边界构成的超平面 (边界超平面). 对于每个 k, $\mathrm{ri}\, F$ 必包

含在 $\operatorname{int} G_k$ 或者 H_k 中. 设 D 是 G_k 或者 H_k 中包含 $\operatorname{ri} F$ 的一些 (当然是有穷个) 相对开凸集之交. 于是 D 是 M 的凸子集，并且也是相对开的. 依据定理 1.9.3, D 是 M 的一个极大相对开凸子集，因此， $\operatorname{ri} F = D$. 但上述这种形式的集合 D 仅有有穷多个，并且 M 的不同面的相对内部不相交 (定理 1.9.2), 因此 M 只可能有有穷多个面.

(2)$\Longrightarrow$(3): 先考察 M 不含任何直线的情形. 根据定理 1.9.7, M 是其极点和极方向的凸包. 由于 M 仅有有穷多个面，故它仅有有穷多个极点和极方向. 因此， M 是有限生成的. 现在假定 M 含有直线，于是 $M = S_0 + L$, 其中 L 是 M 的直线状空间，而 $M_0 = M \cap L^{\perp}$ 是不含直线的闭凸集. 注意 M_0 的面都能表示成 $F_0 = F \cap L^{\perp}$, 其中 F 是 M 的面. 于是 M_0 仅有有穷多个面. 因此， M_0 是有限生成的. 设 $b_1, \cdots, b_m$ 是 L 的基, 任意向量 $x \in M$ 都可以表示成

$$x = x_0 + \lambda_1 b_1 + \cdots + \lambda_m b_m + \lambda_1'(-b_1) + \cdots + \lambda_m'(-b_m),$$

其中 $x_0 \in M_0, \lambda_i \geq 0, \lambda_i' \geq 0, i = 1, \cdots, m$. 这样我们证明了 M 本身也是有限生成的.

(3)$\Longrightarrow$(2): 设 $M = \operatorname{conv} S$, S 是由 $\mathbb{R}^n$ 中的一些点和方向组成的集合. M 的闭性容易直接验证. 此外，依据定理 1.9.4, M 的面与 S 的一些子集之间有着一一对应的关系，从而 M 只可能有有穷多个面.

(2)$\Longrightarrow$(1): 不失一般性，我们只要讨论 $\dim M = n$ 的情形就可以了. 依据定理 3.5.1, M 是其所有切半空间之交. 如果 H 是 M 的一个切半空间的边界超平面，根据定义，存在 $x \in M$, 使得 H 是 M 的通过点 x 的唯一的承托超平面. 因此 H 也是 M 的通过暴露面 $M \cap H$ 的唯一的承托

超平面. 由于 M 仅有有穷多个面, 它也只可能有有穷多个切闭半空间. 因此 M 是一个多面凸集. ▮

推论 3.5.3 $\mathbb{R}^n$ 中的每个多面凸集至多有有穷多个极点和极方向.

证明 这是因为极点和极方向对应于单点和半直线面, 而根据定理 3.5.2, 每个多面凸集的面数是有穷的. ▮

下一个定理在研究线性规划问题时是有用的.

定理 3.5.4 设 P 是 $\mathbb{R}^n$ 中不含任何直线的多面凸集. 假定线性泛函 $f:\mathbb{R}^n\to\mathbb{R}$ 在 P 上有上界, 即, $\sup\{f(x)\mid x\in P\}<\infty$, 那么 P 至少有一个极点 z, 使得 f 在 z 处达到其在 P 上的最大值, 即

$$f(z)=\sup\{f(x)\mid x\in P\}.$$

证明 不失一般性, 我们可以假定 $\dim P=n$. 我们仍采用关于空间维数 n 的归纳法来证明. 对于 $n=1$, 不含任何直线的 "多面凸集" P 是闭线段或者闭半射线, 所要的结论不难直接验证. 假定定理对于维数小于 n 的空间成立. 现在设 P 是 $\mathbb{R}^n$ 中不含任何直线的多面凸集. 对于 $\mathbb{R}^n$ 上的线性泛函 f, 容易看出

$$\mu=\sup\{f(x)\mid x\in P\}=\sup\{f(x)\mid x\in \operatorname{bd}P\}.$$

设 $H_1,\cdots,H_m$ 是 $\mathbb{R}^n$ 中的 m 个超平面使得 $P=\bigcap\limits_{i=1}^{m}H_i^+$, 其中 H_i^+ 和 H_i^- 是与 H_i 相关联的两个闭半空间. 于是

$$\operatorname{bd}P=\bigcup_{j=1}^{m}(H_j\cap P).$$

这样, 存在某个标号 j, $1\le j\le m$, 使得

$$\mu=\sup\{f(x)\mid x\in P\cap H_j\}.$$

注意， $H_j \cap P$ 仍然是一个多面凸集，因为 H_j 可以写成 $H_j = H_j^- \cap H_j^+$. 从而

$$H_j \cap P = \Big(\bigcap_{i=1}^{m} H_i^+ \Big) \cap H_j^- .$$

显然， $\dim(H_j \cap P) < n$, 并且 $H_j \cap P$ 也不含任何直线. 于是由归纳假设知，存在 $P \cap H_j$ 的一个极点 z, 使得 $\mu = f(z)$. 但 $P \cap H_j$ 的极点也是 P 的极点，定理得证. ∎

最后我们简单提一下多面凸函数的概念. $\mathbb{R}^n$ 中的一个凸函数 f 叫做多面凸函数，是指其上图 $\operatorname{epi} f$ 是 $\mathbb{R}^{n+1}$ 中的一个多面凸集. 例如，仿射函数，多面凸集的示性函数就是多面凸函数的例子.

于是对于 $\mathbb{R}^n$ 上的多面凸函数 f, $\operatorname{epi} f$ 必定是 $\mathbb{R}^{n+1}$ 中有穷多个闭半空间之交，这些半空间或者是垂直的，或者是仿射函数的上图. 换句话说，f 是一个多面凸函数，当且仅当 f 可以表示成

$$f(x) = g(x) + \delta(x|M), \quad \forall\, x \in \mathbb{R}^n,$$

其中

$$g(x) = \max\{\langle x, b_i \rangle - \beta_i \mid 1 \le i \le k\},$$

$$M = \{x \in \mathbb{R}^n \mid \langle x, b_j \rangle \le \beta_j,\ k+1 \le j \le m\}.$$

这里 $b_1, \cdots, b_k, b_{k+1}, \cdots, b_m$ 和 $\beta_1, \cdots, \beta_m$ 为给定一些向量和数.

$\mathbb{R}^n$ 上的一个凸函数 f 叫做是有限生成的，是指存在向量 $a_1, \cdots, a_m \in \mathbb{R}^n$ 和标量 $\alpha_1, \cdots, \alpha_m \in \mathbb{R}$, 使得

$$f(x) = \inf\{\lambda_1 \alpha_1 + \cdots + \lambda_m \alpha_m\}, \tag{3.5.1}$$

其中下确界是对满足下列条件的一切标量 λ_i 取的：

$$\lambda_1 a_1 + \cdots + \lambda_k a_k + \lambda_{k-1} a_{k+1} + \cdots + \lambda_m a_m = x,$$

$$\lambda_1 + \cdots + \lambda_k = 1,\ \lambda_i \geq 0, 1 \leq i \leq m.$$

不难看出，由式 (3.5.1) 定义的函数 f 可以表示成

$$f(x) = \inf\{\mu \mid (x, \mu) \in F\},$$

其中 F 是 $\mathbb{R}^{n+1}$ 中由有穷个点 $(a_1, \alpha_1), \cdots, (a_k, \alpha_k)$ 和方向 $(a_{k+1}, \alpha_{k+1}), \cdots, (a_m, \alpha_m)$ 组成的集合的凸包. 根据定理 3.6.2, F 是一闭子集，从而与 $\operatorname{epi} f$ 相等. 这表明，对于使得 $f(x)$ 有穷的任意 $x \in \mathbb{R}^n$, 点 $(x, f(x))$ 必属于 F, 因此式 (3.5.1) 中定义 $f(x)$ 的下确界实际上被达到. 这样我们证明了：

定理 3.5.5 为了 $\mathbb{R}^n$ 上的一个凸函数 f 是多面凸函数，必须且只须它是有限生成的. 这样的函数如果还是真凸的，则必定是闭的. 在有限生成的凸函数的定义中，如果对于给定的 x, 式 (3.5.1) 中的下确界有穷，则它必定在某些 λ_i 处达到.

$\mathbb{R}$ 上的绝对值 $|x|$ 显然是多面凸的. 更一般地，$\mathbb{R}^n$ 上的函数

$$f(x) = |\xi_1| + \cdots + |\xi_n|, \quad \forall x \in \mathbb{R}^n$$

也是多面凸的，因为它可以表示成形式如

$$f_\varepsilon(x) = \varepsilon_1 \xi_1 + \cdots + \varepsilon_n \xi_n, \quad \varepsilon_j = +1 \text{ 或 } -1$$

的 2^n 个线性函数的点点上确界，这里 $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)$.

另一个常遇到的多面凸函数是

$$f(x) = \max\{|\xi_1|, \cdots, |\xi_n|\}.$$

容易看出，f 是 $2n$ 个形式如

$$f_{\varepsilon_j}(x) = \varepsilon_j \xi_j, \quad \varepsilon_j = +1 \text{ 或 } -1,\ 1 \le j \le n$$

的线性函数的点点上确界.

定理 3.5.6 多面凸函数的共轭仍然是多面凸函数.

证明 设 f 是 $I\!R^n$ 上的多面凸函数. 于是 f 是有限生成的，从而可以用式 (3.5.1) 表达. 由此可得

$$f^*(x^*) = \sup\left\{\sum_{i=1}^m \lambda_i \langle a_i, x^*\rangle - \sum_{i=1}^m \lambda_i \alpha_i\right\}, \quad x \in I\!R^n,$$

这里上确界是对满足

$$\lambda_1 \ge 0, \cdots, \lambda_m \ge 0,\ \lambda_1 + \cdots + \lambda_k = 1$$

的一切 $\lambda_1, \cdots, \lambda_m$ 取的. 容易看出，当

$$\langle a_i, x^*\rangle - \alpha_i \le 0, \quad \forall i = k+1, \cdots, m$$

时有

$$f^*(x^*) = \max\{\langle a_i, x^*\rangle - \alpha_i \mid i = 1, \cdots, k\};$$

否则的话，$f^*(x^*) = +\infty$. 因此 f^* 是多面凸的. ▮

习　题　3.5

3.5.1 试证: $I\!R^n$ 中每个多面凸集 M 都可以表示成 $M = \{x \in I\!R^n \mid Bx \le b\}$, 其中 B 是某个 $m \times n$ 矩阵，$b \in I\!R^m$.

3.5.2 设 B 是 $m \times n$ 矩阵，$b \in I\!R^m$. 试证: $M = \{x \in I\!R^n \mid Bx = b, x \ge 0\}$ 是一个多面凸集，并且至少有一个极点.

3.5.3 试证: 当上题中的 M 为无界时，M 必有极方向.

3.5.4 举例说明定理 3.5.4 中关于 P 的两个限制是必要的.

第四章　凸函数的次微分运算

本章讨论凸分析的重要内容之一 —— 凸函数的微分理论. 众所周知，实函数理论由于引进了微积分概念，其面貌就彻底改变了，尤其是它把本来相当困难的函数极值问题变得一目了然 (就可微函数而言). 对于凸函数来说，一个重要的事实是：单侧方向导数总是存在的. 我们通过凸函数 f 的图像的切平面引入所谓次梯度和次微分的概念，使得凸函数理论和应用更加丰富. 我们在 4.1 节中介绍方向导数、次梯度和次微分的定义，并在许多具体情况下给出次微分的具体形式. 4.2 节讨论次微分的基本性质，而 4.3 节则进一步研究次微分映射的一个重要性质 —— 单调性.

4.1　次梯度和次微分

设 $f:\mathbb{R}^n\to[-\infty,\infty]$ 为任意函数，并设 $x\in\mathbb{R}^n$ 使得 $f(x)$ 取有穷值. 对于 $y\in\mathbb{R}^n$, 如果极限

$$f'_+(x;y)=\lim_{\lambda\downarrow 0}\frac{f(x+\lambda y)-f(x)}{\lambda}$$

和

$$f'_-(x;y)=\lim_{\lambda\uparrow 0}\frac{f(x+\lambda y)-f(x)}{\lambda}$$

存在 (有穷值或无穷值都容许), 则其极限 $f'_+(x;y)$ 和 $f'_-(x;y)$ 分别称为 f 在 x 处沿 y 方向的右导数和左导数. 根据定义，显然有

$$f'_-(x;y) = -f'_+(x;-y).$$

如果

$$f'_-(x;y) = f'_+(x;y),$$

即极限

$$f'(x;y) = \lim_{\lambda \to 0} \frac{f(x+\lambda y) - f(x)}{\lambda}$$

存在，则称极限 $f'(x;y)$ 为 f 在 x 处沿 y 方向的方向导数.

自然，若 f 在 x 处可微，则方向导数 $f'(x;y)$ 对一切 y 存在且有穷，同时

$$f'(x;y) = \langle \nabla f(x), y \rangle, \quad \forall\, y \in I\!R^n,$$

其中 $\nabla f(x)$ 表示 f 在 x 处的梯度.

定理 4.1.1 设 $f : I\!R^n \to [-\infty, \infty]$ 为凸函数，并设 $x \in I\!R^n$ 使得 $f(x)$ 为有穷值，那么对于每个 $y \in I\!R^n$, 差商

$$\varphi(\lambda) = \frac{f(x+\lambda y) - f(x)}{\lambda}$$

是 $\lambda > 0$ 的非减函数. 因此 $f'_+(x;y)$ 存在，并且

$$f'_+(x;y) = \inf_{\lambda > 0} \frac{f(x+\lambda y) - f(x)}{\lambda}.$$

此外， $f'_+(x;y)$ 是 $y \in I\!R^n$ 的正齐性凸函数，满足

$$f'_-(x;y) \le f'_+(x;y), \quad \forall\, y \in I\!R^n.$$

证明 由于 f 是凸函数，如果记

$$g(\lambda) = f(x+\lambda y), \quad \lambda \in I\!R^n,$$

则显然 $g : I\!R \to [-\infty, \infty]$ 也是凸函数. 于是利用 2.1 节的办法，对于广义实值凸函数 g, 同样可以证明函数 $\varphi(\lambda) =$

$(g(\lambda)-g(0))/\lambda$ 是 $\lambda>0$ 的非减函数，从而 $f'_+(x;y)$ 和 $f'_-(x;y)$ 存在. $f'_+(x;0)=0$ 是显然的. 剩下证明 $f'_+(x;y)$ 关于 $y\in I\!R^n$ 的正齐性. 设 $y\in I\!R^n$ 和 $\lambda>0$, 那么根据定义,

$$\begin{aligned} f'_+(x;\lambda y) &= \lim_{t\downarrow 0}\frac{f(x+t\lambda y)-f(x)}{t} \\ &= \lim_{t\downarrow 0}\lambda\frac{f(x+t\lambda y)-f(x)}{t\lambda} \\ &= \lambda f'_+(x;y). \end{aligned}$$

最后设 $y,z\in I\!R^n$, 依据定义并利用 f 的凸性得到

$$\begin{aligned} f'_+(x;y+z) &= \lim_{t\downarrow 0}\frac{f(x+ty+tz)-f(x)}{t} \\ &\le \lim_{t\downarrow 0}\frac{f(x+2ty)+f(x+2tz)-2f(x)}{2t} \\ &= \lim_{t\downarrow 0}\frac{f(x+2ty)-2f(x)}{2t} \\ &\quad +\lim_{t\downarrow 0}\frac{f(x+2tz)-2f(x)}{2t} \\ &= f'_+(x;y)+f'_+(x;z). \end{aligned}$$

∎

设 $f: I\!R^n\to[-\infty,\infty]$ 为凸函数. 称向量 $x^*\in I\!R^n$ 为 f 在点 x 处的一个次梯度，是指它满足

$$f(z)\ge f(x)+\langle x^*,z-x\rangle,\quad \forall\, z\in I\!R^n. \tag{4.1.1}$$

这个条件称作次梯度不等式. 当 f 在 x 处存在次梯度 x^*, 并且 $f(x)$ 为有穷值时，必有 $f(z)>-\infty,\forall\, z\in I\!R^n$, 这时它有简单的几何意义：仿射函数

$$h(z)=f(x)+\langle z-x,x^*\rangle$$

的图像 $\{(z, h(z)) \mid z \in \mathbb{R}^n\}$ 是凸集 $\operatorname{epi} f$ 在点 $(x, f(x))$ 处的一个非垂直的承托超平面.

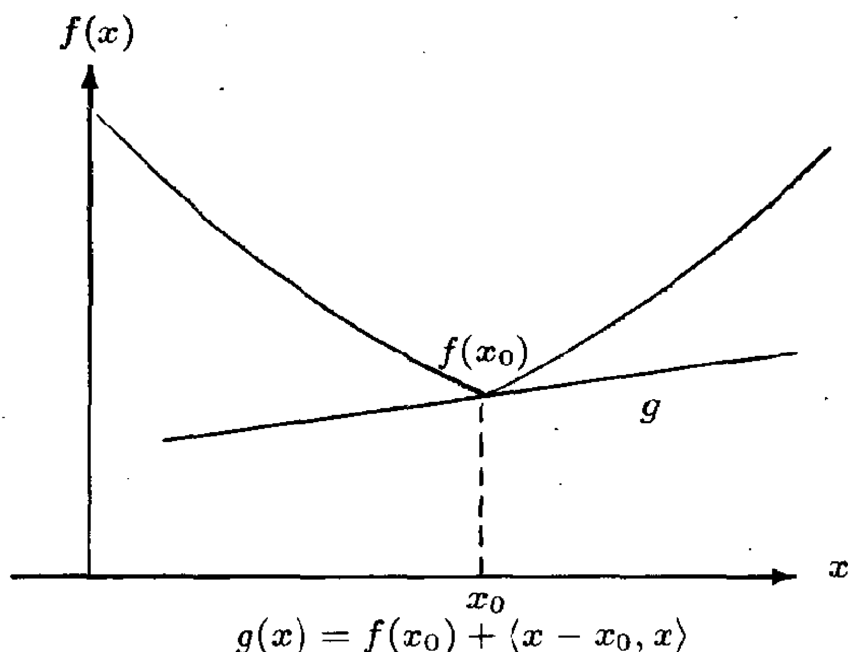


图 4.1.1 函数 f 的次梯度

f 在 x 处的所有次梯度的集合叫做 f 在 x 的次微分，记作 $\partial f(x)$, 即

$$\partial f(x) = \{x^* \in \mathbb{R}^n \mid f(z) \geq f(x) + \langle z - x, x^* \rangle, \forall z \in \mathbb{R}^n\}.$$

这样，一般说来，$\partial f : x \mapsto \partial f(x)$ 是一个多值映射，称为 f 的次微分映射. 不难看出，$\partial f(x)$ 是 $\mathbb{R}^n$ 中的一个闭凸集，因为由定义，$x^* \in \partial f(x)$ 恰好是一族线性不等式的解. 当然对某些 x, $\partial f(x)$ 可以是空集，也可以恰好含有一个向量. 如果 $\partial f(x)$ 不空，则称 f 在 x 处是次可微的.

定理 4.1.2 设 $f : \mathbb{R}^n \to [-\infty, \infty]$ 为凸函数，并设 $x \in \mathbb{R}^n$ 使得 $f(x)$ 为有穷值. 那么 $x^* \in \partial f(x)$, 当且仅当

$$f'_+(x; y) \geq \langle x^*, y \rangle, \quad \forall y \in \mathbb{R}^n. \qquad (4.1.2)$$

此外，$\operatorname{cl} f'_+(x, \cdot)$ 是闭凸集 $\partial f(x)$ 的承托函数.

证明 令 $z = x + \lambda y$, 则次梯度不等式 (4.1.1) 变成

$$[f(x+\lambda y) - f(x)]/\lambda \geq \langle x^*, y\rangle, \quad \forall \lambda > 0, \forall y \in \mathbb{R}^n.$$

由于差商 $[f(x+\lambda y) - f(x)]/\lambda$ 当 $\lambda \downarrow 0$ 时递减收敛于 $f'_+(x;y)$, 故上述不等式与式 (4.1.2) 等价. 最后一个结论是推论 3.2.8 的直接结果. ∎

定理 4.1.3 设 $f: \mathbb{R}^n \to (-\infty, \infty]$ 为真凸函数.

(1) 如果 f 在 x 处次可微, 则 $f(x)$ 是有穷值;

(2) 如果 $x \notin \operatorname{dom} f$, 则 $\partial f(x) = \emptyset$;

(3) 如果 $x \in \operatorname{ri}(\operatorname{dom} f)$, 则 $\partial f(x) \neq \emptyset$, 这时 $f'_+(x;y)$ 作为 y 的函数是闭真凸的, 并且

$$f'_+(x;y) = \sup\{\langle x^*, y\rangle \mid x^* \in \partial f(x)\} = \sigma(y|\partial f(x)), \tag{4.1.3}$$

即 $f'_+(x;y)$ 作为 y 的函数是闭凸集 $\partial f(x)$ 的承托函数;

(4) 最后, $\partial f(x)$ 为非空有界集的充分必要条件是 $x \in \operatorname{int}(\operatorname{dom} f)$, 这时, 对于一切 $y \in \mathbb{R}^n$, $f'_+(x;y)$ 是有穷的.

证明 定理的前两个结论 (1) 和 (2) 是显然的. 今设 $x \in \operatorname{ri}(\operatorname{dom} f)$. 从定理 2.4.13 的证明知道, 存在一个仿射函数 $a(z) = \langle z, x^*\rangle + \alpha$ 使得 $a(x) = f(x)$, 并且

$$f(z) \geq a(z), \quad \forall z \in \mathbb{R}^n.$$

但从 $f(x) = a(x)$ 推出 $a(z) = \langle z - x, x^*\rangle + f(x)$, 从而

$$f(z) \geq f(x) + \langle x^*, z - x\rangle, \quad \forall z \in \mathbb{R}^n,$$

即 $x^* \in \partial f(x)$. 因此 $\partial f(x) \neq \emptyset$.

注意, 当 $x \in \operatorname{ri}(\operatorname{dom} f)$ 时, $f'_+(x;\cdot)$ 的有效定义域是一个仿射集, 即与 $\operatorname{dom} f$ 的仿射包 $\operatorname{aff}(\operatorname{dom} f)$ 相平行的一个子空间. 但 $f'_+(x;0) = 0$, 故 $f'_+(x;0)$ 在此子空间上不可

能恒等于 $-\infty$. 从而依据定理 2.2.11 和推论 2.4.10, $f'_+(x;\cdot)$ 是闭的真凸函数. 从定理 4.1.2 和推论 3.2.8 可知, $f'_+(x;\cdot)$ 本身是 $\partial f(x)$ 的承托函数, 即式 (4.1.3) 成立. 这样 (3) 得证.

现在设 $x\in \mathrm{int}\,(\mathrm{dom}\, f)$, 则 $f'_+(x;\cdot)$ 的有效定义域是全空间 $\mathbb{R}^n$, 故 $\partial f(x)$ 的承托函数 $\sigma(\cdot|\partial f(x))$ 处处有穷. 但依据 (3), $f'_+(x;\cdot)$ 是闭的, 因此 $f'_+(x;\cdot)$ 处处有穷. 根据推论 3.2.9, 非空凸集有界的充分必要条件是其承托函数处处有穷, 因此 $\partial f(x)$ 是有界的. 反之, 若 $\partial f(x)$ 是非空有界的, 则其承托函数 $f'_+(x;\cdot)$ 处处有穷. 从而由推论 1.4.12, $x\in \mathrm{int}\,(\mathrm{dom}\, f)$. 于是定理的最后一个结论 (4) 成立. ∎

这里我们指出两种特别重要的情形. 其一是 f 为 $\mathbb{R}^n$ 上有穷值凸函数, 这时 f 在每一点 $x\in\mathbb{R}^n$ 次可微, 并且 $\partial f(x)$ 是非空有界闭凸集. 其二是 $\partial f(x)$ 由单点 x^* 组成, 这时从定理 4.1.3 得知 $x\in \mathrm{int}\,(\mathrm{dom}\, f)$. 我们将进一步证明, 为了 $\partial f(x)=\{x^*\}$, 必须且只须:

(1) $x\in \mathrm{int}\,(\mathrm{dom}\, f)$;

(2) f 在通常意义下可微, 并且 $x^*=\nabla f(x)$, $\nabla f(x)$ 表示 f 在 x 处的梯度

$$\nabla f(x)=\left(\frac{\partial f}{\partial \xi_1}(x),\cdots,\frac{\partial f}{\partial \xi_n}(x)\right)^{\tau}.$$

我们先回忆一下函数 f 在普通意义下可微的概念. 我们说 f 在 x 处可微, 是指存在一向量 $x^*\in\mathbb{R}^n$(必唯一) 使得

$$f(z)=f(x)+\langle z-x,x^*\rangle+o(\|z-x\|),$$

即

$$\lim_{z\to x}\frac{f(z)-f(x)-\langle z-x,x^*\rangle}{\|z-x\|}=0.$$

这时称 x^* 为 f 在 x 处的梯度，记作 $x^* = \nabla f(x)$. 注意 f 在 x 处可微的定义隐含着 $x \in \operatorname{int}(\operatorname{dom} f)$.

如果用 e_j 表示 $n \times n$ 单位矩阵的第 j 列构成的列向量，则容易看出

$$\langle \nabla f(x), e_j \rangle = \lim_{\lambda \to 0} \frac{f(x + \lambda e_j) - f(x)}{\lambda} = \frac{\partial f}{\partial \xi_j}(x),$$

并且对于任意 $y \in I\!R^n$, $y \neq 0$,

$$\begin{aligned} 0 &= \lim_{\lambda \downarrow 0} \frac{f(x + \lambda y) - f(x) - \langle \nabla f(x), \lambda y \rangle}{\lambda \|y\|} \\ &= [f'_+(x; y) - \langle \nabla f(x), y \rangle] / \|y\|. \end{aligned}$$

因此，$f'_+(x; y)$ 存在并且是 y 的线性函数：

$$f'_+(x; y) = \langle \nabla f(x), y \rangle, \quad \forall y \in I\!R^n.$$

定理 4.1.4 设 $f: I\!R^n \to (-\infty, \infty]$ 为真凸函数，$x \in \operatorname{dom} f$. 如果 f 在 x 处可微，则 $\nabla f(x)$ 是 f 在 x 处的唯一的次梯度，特别地有

$$f(z) \geq f(x) + \langle z - x, \nabla f(x) \rangle, \quad \forall z \in I\!R^n.$$

反之，若 $\partial f(x) = \{x^*\}$，则 f 在 x 处可微，并且 $x^* = \nabla f(x)$.

证明 先假定 f 在 x 处可微，于是

$$f'_+(x; y) = \langle \nabla f(x), y \rangle.$$

由定理 4.1.2, $x^* \in \partial f(x)$ 的充分必要条件是

$$\langle \nabla f(x), y \rangle \geq \langle x^*, y \rangle, \quad \forall y \in I\!R^n.$$

但显然，上式当且仅当 $x^* = \nabla f(x)$ 时成立，于是 $\nabla f(x)$ 是 f 在 x 处的唯一的次梯度.

现在假设 f 在 x 处有唯一的次梯度 x^*, 即 $\partial f(x) = \{x^*\}$. 从定理 4.1.3 知 $x \in \operatorname{int}(\operatorname{dom} f)$. 定义凸函数

$$g(y) = f(x+y) - f(x) - \langle y, x^* \rangle, \quad y \in \mathbb{R}^n.$$

如果 $y^* \in \partial g(0)$, 则由于 $g(0) = 0$,

$$g(y) \geq \langle y, y^* \rangle, \quad \forall\, y \in \mathbb{R}^n,$$

即

$$f(x+y) \geq f(x) + \langle y, x^* + y^* \rangle, \quad \forall\, y \in \mathbb{R}^n.$$

这表明 $x^* + y^* \in \partial f(x)$. 但 $\partial f(x) = \{x^*\}$, 因此 $y^* = 0$. 这样我们证明了 $\partial g(0) = \{0\}$.

为了证明 $x^* = \nabla f(x)$, 只须证明:

$$\lim_{y \to 0} g(y)/\|y\| = 0.$$

注意, $0 \in \operatorname{int}(\operatorname{dom} g)$, 因此 $g'_+(0;\cdot)$ 是闭的. 从而 $g'_+(0;\cdot)$ 是 $\partial g(0)$ 的承托函数. 但 $\partial g(0) = \{0\}$, 故 $g'_+(0;\cdot)$ 恒等于 0. 因此,

$$0 = g'_+(0;z) = \lim_{\lambda \downarrow 0}[g(\lambda z) - g(0)]/\lambda, \quad \forall\, z \in \mathbb{R}^n,$$

这里 $g(0) = 0$, 并且差商 $g(\lambda z)/\lambda$ 是 $\lambda > 0$ 的不减函数. 记

$$h_\lambda(z) = g(\lambda z)/\lambda, \quad \lambda > 0,\ z \in \mathbb{R}^n,$$

则 $h_\lambda(z)$ 当 $\lambda \downarrow 0$ 时点点递减趋于常值函数 0. 现在设 $\overline{B}(0,1)$ 是 $\mathbb{R}^n$ 中的闭单位球, 并设 $\{a_1, \cdots, a_m\}$ 为 $\mathbb{R}^n$ 中的一有限点集, 使得 $\operatorname{conv}\{a_1, \cdots, a_m\} \supset \overline{B}(0,1)$. 于是任意 $z \in \overline{B}(0,1)$ 可以表示成凸组合

$$z = \lambda_1 a_1 + \cdots + \lambda_m a_m, \quad \lambda_1 + \cdots + \lambda_m = 1.$$

利用 h_λ 的凸性得到

$$0 \le h_\lambda(z) \le \sum_{i=1}^{m} \lambda_i h_\lambda(a_i) \le \max\{h_\lambda(a_i) \mid i = 1, \cdots, m\}.$$

但是我们知道，对于每一个 i, $h_\lambda(a_i)$ 当 $\lambda \downarrow 0$ 时递减趋于 0, 因此， $h_\lambda(z)$ 当 $\lambda \downarrow 0$ 时对 $z \in \overline{B}(0,1)$ 一致地趋于 0, 即任给 $\varepsilon > 0$, 存在 $\delta > 0$, 使得

$$0 \ge g(\lambda z)/\lambda \le \varepsilon, \quad \forall \lambda \in (0, \delta], \forall z \in \overline{B}(0,1).$$

但每一向量 $y \in \overline{B}(0,\delta)\backslash\{0\}$ 均可表示成 $y = \lambda z$, 其中 $z = y/\|y\|$, $\lambda = \|y\|$, 因此

$$0 \le g(y)/\|y\| \le \varepsilon, \quad \forall y, 0 < \|y\| \le \delta. \qquad \blacksquare$$

下面我们列举一些常用的重要凸函数的次微分的例子.

例 4.1.1 $f(x) = \|x\|^2/2$, $x \in \mathbb{R}^n$. 我们知道， f 是 $\mathbb{R}^n$ 上的可微函数，并且 $\nabla f(x) = x$, 故

$$\partial f(x) = \{x\}, \quad \forall x \in \mathbb{R}^n.$$

例 4.1.2 $f(x) = \|x\|$. f 在每一点 $x \ne 0$ 处可微，并且 $\nabla f(x) = x/\|x\|$. 为了考虑 f 在 $x = 0$ 的次可微性，由定义， $x^* \in \partial f(0)$, 当且仅当

$$\|z\| \ge \langle z, x^* \rangle, \quad \forall z \in \mathbb{R}^n,$$

即 $\partial f(0) = \overline{B}(0,1)$. 于是

$$\partial f(x) = \begin{cases} \overline{B}(0,1), & x = 0, \\ x/\|x\|, & x \ne 0. \end{cases}$$

例 4.1.3 对于某个 j, $1 \le j \le n$, $f_j(x) = |\xi_j|$, $x = (\xi_1, \cdots, \xi_n)^T \in \mathbb{R}^n$. 用 e_j 表示 $n \times n$ 单位矩阵的第 j 列构

成的列向量，则

$$\partial f_j(x) = \begin{cases} \operatorname{conv}\{\pm e_j\}, & \xi_j = 0, \\ \xi_j e_j/|\xi_j|, & \xi_j \neq 0. \end{cases}$$

为了讨论凸函数 $f(x) = \max\{|\xi_1|, |\xi_2|, \cdots, |\xi_n|\}$, 我们先给出这一类凸函数次微分表达式的一个一般的定理.

定理 4.1.5 设 $f_j : \mathbb{R}^n \to \mathbb{R}$ 为凸函数, $j = 1, \cdots, m$. 令

$$f(x) = \max_{1 \leq j \leq m} f_j(x), \quad x \in \mathbb{R}^n.$$

那么

$$\partial f(x) = \left\{ \sum_{i \in I(x)} \lambda_i x_i^* \,\middle|\, \sum_{i \in I(x)} \lambda_i = 1,\ \lambda_i \geq 0,\ x_i^* \in \partial f_i(x) \right\},$$

其中

$$I(x) = \{i \mid 1 \leq i \leq m, f_i(x) = f(x)\}.$$

证明 我们知道, $f : \mathbb{R}^n \to \mathbb{R}$ 是凸函数. 如果记

$$M(x) = \left\{ \sum_{i \in I(x)} \lambda_i x_i^* \,\middle|\, \sum_{i \in I(x)} \lambda_i = 1,\ \lambda_i \geq 0,\ x_i^* \in \partial f_i(x) \right\},$$

则由于每个 $\partial f_i(x)$ 是非空有界闭凸集 (见定理 4.1.3), $M(x)$ 也是 $\mathbb{R}^n$ 上非空有界闭凸集.

任取 $y \in \mathbb{R}^n$, 对于每个 $\lambda > 0$, 用 $i(\lambda)$ 表示 $I(x + \lambda y)$ 中的任意一个指标. 由于 $I(x+\lambda y) \subset \{1, \cdots, m\}$, 总能找到一列正数 $\lambda_j \downarrow 0\ (j \to \infty)$, 使得对于一切 j, $i(\lambda_j) = k$. 今证 $k \in I(x)$. 事实上, 若 $k \notin I(x)$, 则 $f_k(x) < f(x)$, 并且由于 $f(x+\lambda y)$ 和 $f_k(x+\lambda y)$ 相对于 λ 连续, 故当 λ_j 充分小时有

$$f_k(x+\lambda_j y) < f(x+\lambda_j y),$$

这与 $k \in I(x+\lambda_j y)$ 矛盾，所以 $k \in I(x)$.

由关系式

$$[f(x+\lambda_j y) - f(x)]/\lambda_j = [f_k(x+\lambda_j y) - f_k(x)]/\lambda_j$$

可见 $f'_+(x;y) = f'_{k+}(x;y)$. 另一方面，对任意 $i \in I(x)$, $\lambda > 0$, 有

$$[f(x+\lambda y) - f(x)]/\lambda \geq [f_i(x+\lambda y) - f_i(x)]/\lambda.$$

因此 $f'_+(x;y) \geq f'_{i+}(x;y)$, $\forall i \in I(x)$. 由此可知

$$f'_+(x;y) = \max\{\langle y, x^*\rangle \mid x^* \in \partial f(x)\} = \max_{i\in I(x)} f'_{i+}(x;y).$$

但是

$$\begin{aligned}
&\max_{i\in I(x)} f'_{i+}(x;y)\\
&= \max\left\{\sum_{i\in I(x)} \lambda_i f'_{i+}(x;y) \,\middle|\, \sum_{i\in I(x)} \lambda_i = 1, \lambda_i \geq 0\right\}\\
&= \max_{\sum_{i\in I(x)}\lambda_i=1,\lambda_i\geq 0} \sum_{i\in I(x)} \lambda_i \max_{x_i^*\in\partial f_i(x)} \langle y, x_i^*\rangle\\
&= \max_{\sum_{i\in I(x)}\lambda_i=1,\lambda_i\geq 0} \max_{x_i^*\in\partial f_i(x)} \left\langle \sum_{i\in I(x)} \lambda_i x_i^*, y\right\rangle\\
&= \max\{\langle y, x^*\rangle \mid x^* \in M(x)\}.
\end{aligned}$$

因此

$$\sigma(y|\partial f(x)) = \sigma(y|M(x)), \quad \forall y \in \mathbb{R}^n.$$

于是由推论 3.2.5, 两个闭凸集 $\partial f(x)$ 和 $M(x)$ 相等.

例 4.1.4 设

$$f(x)=\max\{|\xi_1|,|\xi_2|,\cdots,|\xi_n|\},\ x\in(\xi_1,\cdots,\xi_n)^\tau\in I\!R^n.$$

由定理 4.1.5 和例 4.1.3 直接可得

$$\partial f(x)=\begin{cases}\operatorname{conv}\{\pm e_1,\cdots,\pm e_n\}, & x=0,\\ \operatorname{conv}\{\xi_j e_j/|\xi_j| \mid j\in I(x)\}, & x\neq 0,\end{cases}$$

其中

$$I(x)=\{j\mid 1\le j\le m,\ f(x)=|\xi_j|\}.$$

对于一个真凸函数 f, 从上面的讨论我们知道,

$$\operatorname{ri}(\operatorname{dom} f)\subset\operatorname{dom}(\partial f)\subset\operatorname{dom} f,$$

其中 $\operatorname{dom}(\partial f)=\{x\in I\!R^n\mid\partial f(x)\neq\emptyset\}$ 为 ∂f 的有效定义域.

例 4.1.5 考虑 $I\!R^n$ 上的凸函数 (见图 4.1.2)

$$f(x)=\begin{cases}-(1-\|x\|^2)^{1/2}, & \|x\|\le 1,\\ +\infty, & \|x\|>1.\end{cases}$$

当 $\|x\|<1$ 时 $f(x)$ 是可微的, 并且 $\nabla f(x)=x/(1-\|x\|^2)^{1/2}$, 而当 $\|x\|>1$ 时 $\partial f(x)=\emptyset$. 如果对于某个 $x\in I\!R^n,\|x\|=1$, 存在次梯度 $x^*\in\partial f(x)$. 于是

$$-(1-\|z\|^2)^{1/2}\ge\langle z-x,x^*\rangle,\quad\forall\, z\in I\!R^n,\|z\|\le 1.$$

特别是, 当取 $z=\alpha x,0<\alpha<1$ 时, 就得到 $(1+\alpha)^{1/2}(1-\alpha)^{-1/2}\le\langle x,x^*\rangle,\forall\,0<\alpha<1$. 这是不可能的. 因此

$$\partial f(x)=\begin{cases}\{x/(1-\|x\|^2)^{1/2}\}, & \|x\|<1,\\ \emptyset, & \|x\|\ge 1.\end{cases}$$

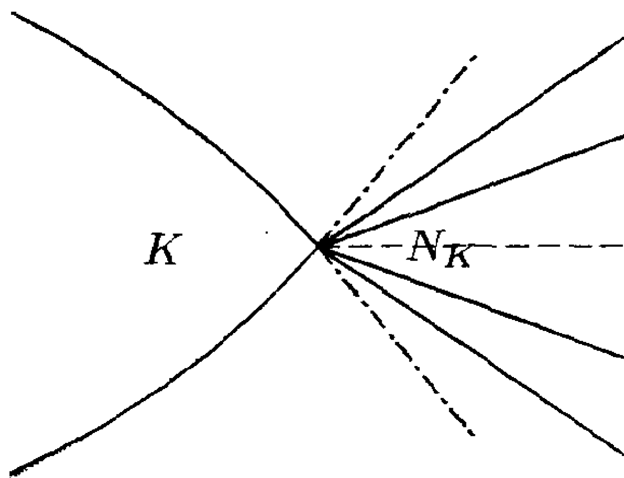


图 4.1.2　法向锥

例 4.1.6　设 $K \subset \mathbb{R}^n$ 为非空凸集，考察 K 的示性函数 $\delta(x|K)$. 由定义，$x^* \in \partial\delta(x|K)$ 当且仅当

$$\delta(z|K) \geq \delta(x|K) + \langle z - x, x^* \rangle, \quad \forall z \in \mathbb{R}^n.$$

这个条件意味着 $x \in K$, 并且

$$\langle z - x, x^* \rangle \leq 0, \quad \forall z \in K. \tag{4.1.4}$$

由式 (4.1.3) 决定的所有 x^* 构成 $\mathbb{R}^n$ 中的一个闭凸锥，称为 K 在 x 处的法向锥，记作 $N_K(x)$. 几何上，$N_K(x)$ 就是与所有向量 $z - x\,(z \in K)$ 的夹角不小于 $\pi/2$ 的向量全体. 当 K 为 $\mathbb{R}^n$ 中的子空间时，对于 $x \in K$,

$$\delta(x|K) = \{x^* \in \mathbb{R}^n \mid \langle z, x^* \rangle = 0, \forall z \in K\},$$

即 $\partial\delta(x|K) = N_K(x)$ 是 K 的直交子空间.

例 4.1.7　$f(x) = \langle x, x^* \rangle + \alpha$ 为 $\mathbb{R}^n$ 上的仿射函数，那么

$$\partial f(x) = \{x^*\}, \quad \forall x \in \mathbb{R}^n.$$

习 题 4.1

4.1.1 给定 $I\!R$ 上的凸函数

$$f(x)=\begin{cases}|x|-2(1-x)^{1/2}, & -3\le x\le 1,\\ +\infty, & \text{其余处}.\end{cases}$$

试证:

$$\partial f(x)=\begin{cases}\emptyset, & x\ge 1,\\ \{1+(1-x)^{-1/2}\}, & 0<x<1,\\ [0,2], & x=0,\\ \{-1+(1-x)^{-1/2}, & -3\le x\le 0,\\ (-\infty,-1/2], & x=-3,\\ \emptyset, & x<-3.\end{cases}$$

4.1.2 设 $\|\cdot\|_p$ 是 $I\!R^n$ 的 p 范数 $(p\ge 1)$, 即

$$\|x\|_p=\left(\sum_{i=1}^{n}|\xi_i|^p\right)^{1/p},\quad \forall\, x=(\xi_1,\cdots,\xi_n)^\tau\in I\!R^n.$$

那么 $f(x)=\|x\|_p$ 是凸函数. 试找出 $f(x)$ 的次微分 $\partial f(x)$.

4.1.3 设

$$f(x)=\begin{cases}-\lambda_1\beta_1-\lambda_2\beta_2, & \text{如果}\, x=\lambda_1 a_1+\lambda_2 a_2,\\ +\infty, & \text{如果}\, x\notin \operatorname{span}\{a_1,a_2\},\end{cases}$$

其中 $a_1,a_2\in I\!R^n$ 为给定的向量, $\beta_1,\beta_2\in I\!R$ 为给定的实数. 试找出 $f(x)$ 的次微分 $\partial f(x)$.

4.1.4 设 $a_1,a_2\in I\!R^n$ 为给定的向量, $\beta_1,\beta_2\in I\!R$ 为给定的实数. 设

$$f(x)=\max\{\langle x,a_1\rangle+\beta_1,\ \langle x,a_2\rangle+\beta_2\},\quad x\in I\!R^n.$$

试证:

$$\partial f(x)=\begin{cases}\{a_1\}, & \text{如果 } \langle x,a_1-a_2\rangle<\beta_2-\beta_1,\\ \{a_2\}, & \text{如果 } \langle x,a_1-a_2\rangle>\beta_2-\beta_1,\\ [a_1,a_2], & \text{如果 } \langle x,a_1-a_2\rangle=\beta_2-\beta_1,\end{cases}$$

这里 $[a_1,a_2]$ 表示连接 a_1 和 a_2 的闭线段.

4.1.5 求出下列凸函数 f 的次微分 ∂f:

(1) $f(x)=\max\{\xi_1,\cdots,\xi_n\},\quad x=(\xi_1,\cdots,\xi_n)^\tau\in I\!R^n$;

(2) $f(x)=|\xi_1|+\cdots+|\xi_n|,\quad x=(\xi_1,\cdots,\xi_n)^\tau\in I\!R^n$;

(3) $f(x)=\begin{cases}0, & \|x\|\le 1,\\ +\infty, & \|x\|>1,\end{cases}\quad x\in I\!R^n.$

4.2 次微分的基本性质

本节讨论凸函数次微分映射的基本性质.

首先, 我们对 $I\!R^n$ 中多值 (也称集值) 映射的概念作一些说明. 我们用 $2^{I\!R^n}$ 表示由 $I\!R^n$ 的全体子集构成的集族. 设 $A: I\!R^n\to 2^{I\!R^m}$ 为一多值映射, 即对于 $x\in I\!R^n$, Ax 是 $I\!R^m$ 中的一个子集. 设 $B: I\!R^n\to 2^{I\!R^m}$ 为另一多值映射, 并设 $\lambda\in I\!R$. 我们约定多值映射的数乘、加法、逆等等运算如下:

(1) $(\lambda A)x=\{\lambda y\mid y\in Ax\}$;

(2) $(A+B)x=Ax+Bx=\{y+z\mid y\in Ax,\ z\in Bx\}$;

(3) $A^{-1}y=\{x\in I\!R^n\mid y\in Ax\}$.

A 的有效定义域 $\operatorname{dom}A$ 和值域 $R(A)$ 分别指

$$\operatorname{dom}A=\{x\in I\!R^n\mid Ax\neq\emptyset\},$$

$$R(A)=\{y\in Ax\mid x\in\operatorname{dom}A\}.$$

从上述规定看到，多值映射 A 的逆 A^{-1} 总是存在的，并且也是一个多值映射 (有可能是一个单值映射).

定理 4.2.1 设 $f: \mathbb{R}^n \to (-\infty, \infty]$ 为真凸函数，$x \in \mathbb{R}^n$, $x^* \in \mathbb{R}^n$. 那么下列四个命题彼此等价:

(1) $x^* \in \partial f(x)$;

(2) $\langle x, x^* \rangle - f(x) = \max\{\langle z, x^* \rangle - f(z) \mid z \in \mathbb{R}^n\}$;

(3) $f(x) + f^*(x^*) \leq \langle x, x^* \rangle$;

(4) $f(x) + f^*(x^*) = \langle x, x^* \rangle$.

此外，如果 $\mathrm{cl}\, f(x) = f(x)$ (f 本身不必是闭的), 则上述彼此等价的命题还可以增加三个:

(5) $x \in \partial f^*(x^*)$;

(6) $\langle x, x^* \rangle - f^*(x^*) = \max\{\langle x, z^* \rangle - f^*(z^*) \mid z^* \in \mathbb{R}^n\}$;

(7) $x^* \in \partial(\mathrm{cl}\, f)(x)$.

证明 当 (1) 成立，即 $x^* \in \partial f(x)$ 时，依据次梯度 x^* 的定义，有

$$\langle z, x^* \rangle - f(z) \leq \langle x, x^* \rangle - f(x), \quad \forall\, z \in \mathbb{R}^n, \tag{4.2.1}$$

这正好是命题 (2). 依照共轭函数的定义，$\langle z, x^* \rangle - f(z)$ 关于 $z \in \mathbb{R}^n$ 的上确界等于 $f^*(x^*)$, 于是从式 (4.2.1) 可知 (2) 又与 (3) 和 (4) 等价. 考虑到共轭函数的性质，(5), (6) 和 (7) 又都等价于

$$f^*(x^*) + f^{**}(x) = \langle x, x^* \rangle.$$

但当 $(\mathrm{cl}\, f)(x) = f(x)$ 时，$f^{**}(x) = (\mathrm{cl}\, f)(x) = f(x)$, 所以上述等式又与 (4) 等价. ▮

推论 4.2.2 设 $f: \mathbb{R}^n \to (-\infty, \infty]$ 为闭的真凸函数，那么 ∂f^* 是 ∂f 的逆映射，即 $\partial f^* = (\partial f)^{-1}$; 换句话

说，$x \in \partial f^*(x^*)$ 当且仅当 $x^* \in \partial f(x)$.

推论 4.2.3 设 $f : \mathbb{R}^n \to (-\infty, \infty]$ 为真凸函数，f 在 x 处次可微, 那么 $(\mathrm{cl}\, f)(x) = f(x)$, 并且 $\partial(\mathrm{cl}\, f)(x) = \partial f(x)$.

证明 事实上，我们有

$$f(x) \geq (\mathrm{cl}\, f)(x) = f^{**}(x) \geq \langle x, x^* \rangle - f^*(x^*).$$

如果 f 在 x 处次可微, 则至少有一向量 x^*, 使得 (4) 成立, 从而 $f(x) = (\mathrm{cl}\, f)(x)$. 然后再利用定理 4.2.1 中的 (1) 和 (7) 的等价性，得到 $\partial(\mathrm{cl}\, f)(x) = \partial f(x)$. ∎

推论 4.2.4 设 $K \subset \mathbb{R}^n$ 为非空闭凸集. 那么对于每一个 $x^* \in \mathbb{R}^n$, $x \in \partial\sigma(x^*|K)$ 等价于线性函数 $\langle \cdot, x^* \rangle$ 在 x 处达到其在 K 上的最大值.

证明 取 $f = \delta(\cdot|K)$, 于是 f^* 是 K 的承托函数 $\sigma(\cdot|K)$. 然后再利用 (5) 和 (2) 的等价性. ∎

推论 4.2.5 设 $f, f_1, f_2 : \mathbb{R}^n \to (-\infty, \infty]$ 为真凸函数，并设 $\lambda \geq 0$. 那么，

(1) $\partial(\lambda f)(x) = \lambda\partial f(x), \quad \forall x \in \mathbb{R}^n$;

(2) $\partial f_1(x) + \partial f_2(x) \subset \partial(f_1 + f_2)(x), \quad \forall x \in \mathbb{R}^n$;

(3) $f(x_0) = \min\{f(x) \mid x \in \mathbb{R}^n\}$, 当且仅当 $0 \in \partial f(x_0)$;

(4) 对于 $\mathbb{R}^n$ 上的仿射函数 $h(x) = \langle x, x^* \rangle + \alpha$, 有

$$\partial(f + h)(x) = \partial f(x) + \partial h(x), \quad \forall x \in \mathbb{R}^n. \tag{4.2.2}$$

证明 (1) 是显然的. 今证 (2). 设 $x_1^* \in \partial f_1(x)$, $x_2^* \in \partial f_2(x)$. 于是

$$f_1(z) \geq f_1(x) + \langle z - x, x^* \rangle, \quad \forall z \in \mathbb{R}^n, \tag{4.2.3}$$

$$f_2(z) \geq f_2(x) + \langle z - x, x^* \rangle, \quad \forall z \in I\!R^n, \tag{4.2.4}$$

把式 (4.2.3) 和式 (4.2.4) 相加，得到

$$(f_1 + f_2)(z) \geq (f_1 + f_2)(x) + \langle z - x, x_1^* + x_1^* \rangle \quad \forall z \in I\!R^n.$$

这表明 $x_1^* + x_2^* \in \partial(f_1 + f_2)(x)$.

为证 (3), 只须注意 $0 \in \partial f(x_0)$ 当且仅当

$$f(z) \geq f(x) + \langle z - x, 0 \rangle = f(x_0), \quad \forall z \in I\!R^n.$$

最后证明 (4). 设 $x^* \in \partial(f + h)(x)$, 于是

$$f(z) + \langle z, x_0^* \rangle + \alpha \geq f(x) + \langle x, x_0^* \rangle + \alpha + \langle z - x, x^* \rangle, \quad \forall z \in I\!R^n,$$

即

$$f(z) \geq f(x) + \langle z - x, x^* - x_0^* \rangle, \quad \forall z \in I\!R^n.$$

根据定义， $x^* - x_0^* \in \partial f(x)$. 但由例 4.1.7, $\partial h(x) = \{x_0^*\rangle$. 因此我们证明了

$$\partial(f + h)(x) \subset \partial f(x) + \partial h(x).$$

由此根据已经证明的 (2), 得知式 (4.2.2) 成立. ∎

下一个定理是次微分运算的基本定理.

定理 4.2.6 (Rockafellar定理) 设 $f_k : I\!R^n \to (-\infty, \infty]$ 为真凸函数， $k = 1, 2, \cdots, m$. 如果 $\mathrm{ri}\,(\mathrm{dom}\, f_1) \cap \cdots \cap \mathrm{ri}\,(\mathrm{dom}\, f_m) \neq \emptyset$, 那么 $\forall x \in I\!R^n$,

$$\partial(f_1 + \cdots + f_m)(x) = \partial f_1(x) + \cdots + \partial f_m(x). \tag{4.2.5}$$

证明 利用数学归纳法，我们只要对 $m = 2$ 证明就够了. 而且依定理 4.2.5, 我们只要证明

$$\partial(f_1 + f_2)(x) \subset \partial f_1(x) + \partial f_2(x).$$

于是设 $x_0^* \in \partial(f_1 + f_2)(x_0)$, 我们要证明 $x_0^* \in \partial f_1(x_0) +$

$\partial f_2(x_0)$. 不失一般性，我们可以假定 $x_0 = x_0^* = 0$, 并且 $f_1(0) = f_2(0) = 0$. 否则令

$$g_1(x) = f_1(x + x_0) - f_1(x_0) - \langle x, x_0^* \rangle,$$

$$g_2(x) = f_2(x + x_0) - f_2(x_0),$$

并且用真凸函数 g_1 和 g_2 代替 f_1 和 f_2 来进行讨论. 于是由定理 4.2.5, 我们有

$$\min\{(f_1 + f_2)(x) \mid x \in \mathbb{R}^n\} = (f_1 + f_2)(0) = 0. \quad (4.2.6)$$

今定义 $\mathbb{R}^{n+1}$ 中的两个凸子集 M_1 和 M_2,

$$M_1 = \operatorname{epi} f_1 = \{(x, \lambda) \in \mathbb{R}^{n+1} \mid f_1(x) \leq \lambda\},$$

$$M_2 = \{(x, \mu) \in \mathbb{R}^{n+1} \mid f_2(x) \leq -\mu\}.$$

由引理 2.4.2 得知

$$\operatorname{ri} M_1 = \{(x, \lambda) \mid x \in \operatorname{ri}(\operatorname{dom} f_1), f_1(x) < \lambda\},$$

$$\operatorname{ri} M_2 = \{(x, \mu) \mid x \in \operatorname{ri}(\operatorname{dom} f_2), f_2(x) < -\mu\}.$$

由于式 (4.2.6), 我们有

$$\operatorname{ri} M_1 \cap \operatorname{ri} M_2 = \emptyset.$$

因此，依定理 1.5.4 及其注， M_1 和 M_2 可以用 $\mathbb{R}^{n+1}$ 中的某个超平面真分离，即存在不同时为 0 的 $x^* \in \mathbb{R}^n$ 和 $\alpha^* \in \mathbb{R}$, 使得

$$\begin{cases} \langle x, x^* \rangle + \lambda\alpha^* < \langle y, x^* \rangle + \mu\alpha^*, \\ \forall f_1(x) < \lambda, \forall f_2(y) < -\mu. \end{cases} \quad (4.2.7)$$

由于式 (4.2.7) 左边的 λ 可以任意大，所以 $\alpha^* \leq 0$. 而如果 $\alpha^* = 0$, 则

$$\langle x, x^* \rangle < \langle y, x^* \rangle, \quad \forall x \in \operatorname{dom} f_1, \forall y \in \operatorname{dom} f_2.$$

但依假设 $\operatorname{ri}(\operatorname{dom} f_1) \cap \operatorname{ri}(\operatorname{dom} f_2) \neq \emptyset$, 故上式不可能成立. 这样我们证明了 $\alpha^* < 0$.

最后, 令 $x_1^* = -x^*/\alpha^*$, 则从式 (4.2.7) 得到

$$-\langle x, x_1^* \rangle + \lambda \geq -\langle y, x_1^* \rangle + \mu, \quad \forall (x, \lambda) \in M_1, \forall (y, \mu) \in M_2.$$

特别取 $\lambda = f_1(x)$, $\mu = -f_2(y)$, $x \in \operatorname{dom} f_1$, $y \in \operatorname{dom} f_2$, 则

$$\begin{cases} f_1(x) - \langle x, x_1^* \rangle \geq -f_2(y) - \langle y, x_1^* \rangle, \\ \forall x \in \operatorname{dom} f_1, \forall y \in \operatorname{dom} f_2. \end{cases} \tag{4.2.8}$$

在式 (4.2.8) 中取 $y = 0$, 则

$$f_1(x) \geq \langle x, x_1^* \rangle = f_1(0) + \langle x - 0, x_1^* \rangle, \quad x \in \operatorname{dom} f_1,$$

故 $x_1^* \in \partial f_1(0)$. 在式 (4.2.8) 中再取 $x = 0$, 则

$$f_2(y) \geq -\langle y, x_1^* \rangle = f_2(0) + \langle y - 0, -x_1^* \rangle, \quad y \in \operatorname{dom} f_2,$$

故 $-x_1^* \in \partial f_2(0)$. 这样, 我们证明了

$$0 = x_1^* + (-x_1^*) \in \partial f_1(0) + \partial f_2(0).$$ ∎

注 定理中的条件 $\operatorname{ri}(\operatorname{dom} f_1) \cap \operatorname{ri}(\operatorname{dom} f_2) \neq \emptyset$ 可减弱为

$$\operatorname{ri}(\operatorname{dom} f_1) \cap \operatorname{dom} f_2 \neq \emptyset,$$

但却不能减弱为

$$\operatorname{dom} f_1 \cap \operatorname{dom} f_2 \neq \emptyset.$$

例如, 考察凸函数 $f_1, f_2 : \mathbb{R} \to (-\infty, \infty]$,

$$f_1(x) = \begin{cases} 0, & \text{当} x < 0, \\ 1, & \text{当} x = 0, \\ +\infty, & \text{当} x > 0; \end{cases}$$

$$f_2(x) = \begin{cases} +\infty, & 当 x < 0, \\ 1, & 当 x = 0, \\ 0, & 当 x > 0. \end{cases}$$

显然 $\operatorname{dom} f_1 = (-\infty, 0]$, $\operatorname{dom} f_2 = [0, +\infty)$, 于是 $\operatorname{dom} f_1 \cap \operatorname{dom} f_2 \neq \emptyset$. 容易算出

$$\partial f_1(x) = \begin{cases} \{0\}, & 当 x < 0, \\ \emptyset, & 当 x \geq 0; \end{cases}$$

$$\partial f_2(x) = \begin{cases} \emptyset, & 当 x \leq 0, \\ \{0\}, & 当 x > 0. \end{cases}$$

但是

$$(f_1 + f_2)(x) = \begin{cases} +\infty, & 当 x \neq 0, \\ 2, & 当 x = 0; \end{cases}$$

$$\partial(f_1 + f_2)(x) = \begin{cases} \emptyset, & 当 x \neq 0, \\ \mathbb{R}, & 当 x = 0. \end{cases}$$

这样我们有

$$\partial(f_1 + f_2)(0) \neq \partial f_1(0) + \partial f_2(0).$$

下一个定理在次梯度计算中也是很有用的.

定理 4.2.7 设 $g : \mathbb{R}^m \to (-\infty, \infty]$ 为真凸函数, A 是 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性变换, 并设 $f(x) = g(Ax), \forall x \in \mathbb{R}^n$. 那么,

$$\partial f(x) \supset A^* \partial g(Ax), \quad \forall x \in \mathbb{R}^n.$$

如果 $R(A) \cap \operatorname{ri}(\operatorname{dom} g) \neq \emptyset$, 这里 $R(A)$ 表示 A 的值域, $R(A) = \{Ax \mid x \in \mathbb{R}^n\}$, 那么

$$\partial f(x) = A^* \partial g(Ax), \quad \forall x \in \mathbb{R}^n.$$

证明 如果 $x^* \in A^*\partial g(Ax)$, 则存在 $y^* \in \partial g(Ax)$ 使得 $x^* = A^*y^*$. 于是对于一切 $z \in \mathbb{R}^n$,

$$f(z) = g(Az) \geq g(Ax) + \langle Az - Ax, y^* \rangle = f(x) + \langle z - x, x^* \rangle,$$

故 $x^* \in \partial f(x)$. 另一方面, 若 $R(A) \cap \text{ri}(\text{dom}\, g) \neq \emptyset$, 则依定理 3.4.6, f 是真凸的, 并且

$$f^*(x^*) = \inf\{g^*(y^*) \mid A^*y^* = x^*\}.$$

这里对于 $f^*(x^*) \neq +\infty$ 的 x^*, 下确界于某个 y^* 处达到, 这时 $A^*y^* = x^*$. 因此对于任意 $x^* \in \partial f(x)$, 存在向量 y^* 使得 $f^*(x^*) = g^*(y^*)$. 同时我们有

$$f(x) + f^*(x^*) = \langle x, x^* \rangle.$$

综上所述, 得到

$$f(x) + g^*(y^*) = \langle x, A^*y^* \rangle,$$

即

$$g(Ax) + g^*(y^*) = \langle Ax, y^* \rangle.$$

由定理 4.2.1 可得 $y^* \in \partial g(Ax)$. 于是 $x^* \in A^*\partial g(Ax)$. ▮

在经典分析中我们知道, 对于一个光滑函数 f, f 在 x 处的梯度与通过该点 x 的切平面垂直. 关于次梯度的这一类似结果则可用凸集的法向锥来描述.

定理 4.2.8 设 f 为 $\mathbb{R}^n$ 上的真凸函数. 假定 f 在 x 处次可微, 但 f 在 x 处并不达到其极小值. 那么, 集合 $M = \{z \in \text{dom}\, f \mid f(z) \leq f(x)\}$ 在 x 处的法向锥是由 $\partial f(x)$ 所生成的闭凸锥.

证明 由于 $f(x)$ 不是 f 的极小值，根据定理 2.4.20, 集合 $N=\{z \mid f(z)<f(x)\}$ 与 M 有相同的闭包. 因此, 为了 x^* 是 M 在 x 处的法向量, 必须且只须

$$\langle z-x, x^*\rangle \leq 0, \quad \forall z \in N.$$

但是依据定理 4.1.1,

$$f'_+(x;y)<0 \Longleftrightarrow y=\lambda(z-x),\, \lambda>0,\, z \in N.$$

因此如果令

$$K=\{y \mid f'_+(x;y)<0\},$$

则 K 是由 $N-x$ 所生成的凸锥. 于是 M 在 x 处的法向锥 K_1 是 K 的极化锥, 即 $K_1=K^o$. 又依定理 2.4.20 和定理 4.1.2,

$$\begin{aligned}
\mathrm{cl}\, K &= \{y \mid \mathrm{cl}_y f'_+(x;y) \leq 0\}=\{y \mid \sigma(y|\partial f(x)) \leq 0\} \\
&= \{y \mid \langle y, x^*\rangle \leq 0, \forall x^* \in \partial f(x)\}=K_2^o,
\end{aligned}$$

其中 K_2 表示由 $\partial f(x)$ 生成的凸锥. 于是

$$K_1=K^o=(\mathrm{cl}\, K)^o=K_2^{oo}=\mathrm{cl}\, K_2,$$

这正是所要证的. ∎

推论 4.2.9 设 f 为 $\mathbb{R}^n$ 上的真凸函数, 并设 $x \in \mathrm{int}\,(\mathrm{dom}\, f)$, $f(x)$ 不是 f 的极小值. 那么为了 x^* 是集合 $M=\{z \mid f(z) \leq f(x)\}$ 在 x 处的法向量, 必须且只须存在一 $\lambda \geq 0$ 使得$x^* \in \lambda \partial f(x)$.

证明 根据定理 4.1.3, 推论的假设意味着 $\partial f(x)$ 是不包含原点的非空有界闭凸集. 显然, 这时 $\partial f(x)$ 所生成的闭凸锥正好是 $\cup\{\lambda \partial f(x) \mid \lambda \geq 0\}$. ∎

习　题　4.2

4.2.1　设 $g(\cdot,\cdot): I\!R^n \times I\!R^m \to (-\infty,+\infty]$ 为真凸函数. 令

$$f(x) = \inf\{g(x,y) \mid y \in I\!R^m\}.$$

如果 $\tilde{y} \in I\!R^m$ 满足 $f(x) = g(x,\tilde{y})$, 试证:

$$x^* \in \partial f(x) \Longleftrightarrow (x^*,0) \in (\partial g)(x,\tilde{y}).$$

4.2.2　设 $M \subset I\!R^n$ 为非空凸集, 我们已经定义了 M 在 x 处的法向锥 $N_M(x) = \partial\delta(\cdot|M)(x)$:

$$N_M(x) = \{x^* \in I\!R^n \mid \langle x^*, y - x\rangle \leq 0, \forall\, y \in M\}.$$

法向锥 $N_M(x)$ 的极化锥 $T_M(x)$ 叫做 M 在 x 处的切向锥, 即

$$T_M(x) = \{x \in I\!R^n \mid \langle x^*, x\rangle \leq 0, \forall\, x^* \in N_M(x)\}.$$

试证:

$$T_M(x) = \mathrm{cl}\left[\bigcup_{\lambda>0} \lambda(M - x)\right].$$

4.2.3　设 g 为 $I\!R^n$ 上的仿射函数, $g(x) = \langle z, x\rangle + \alpha$, 记 $M = \{x \in I\!R^n \mid g(x) \leq 0\}$, 试证:

$$N_M(x) = \{\lambda z \mid \lambda \geq 0\};$$

$$T_M(x) = M - x.$$

4.2.4　设 $M_1, M_2 \subset I\!R^n$ 为两个凸集.

(1) 如果 $M_1 \subset M_2$, 并且 $x \in M_1$, 那么

$$T_{M_1}(x) \subset T_{M_2}(x), \quad N_{M_1}(x) \supset N_{M_2}(x).$$

(2) 如果 $x_1 \in M_1, x_2 \in M_2$, 那么

$$T_{M_1+M_2}(x_1 + x_2) = \mathrm{cl}\left[T_{M_1}(x_1) + T_{M_2}(x_2)\right],$$

$$N_{M_1+M_2}(x_1+x_2)=N_{M_1}(x_1)\cap T_{M_2}(x_2).$$

(3) 如果 $x\in M_1\cap M_2$, 并且 $\operatorname{int} M_1\cap M_2\neq\emptyset$ 或 $\operatorname{ri} M_1\cap \operatorname{ri} M_2\neq\emptyset$, 那么

$$T_{M_1\cap M_2}(x)=T_{M_1}(x)\cap T_{M_2}(x),$$

$$N_{M_1\cap M_2}(x)=N_{M_1}(x)+N_{M_2}(x).$$

4.3 次微分映射的单调性*

在 2.2 节中我们曾指出，对于一个可微的凸函数 $f:\mathbb{R}^n\to\mathbb{R}$, 有

$$\langle\nabla f(x)-\nabla f(y),x-y\rangle\geq 0,\quad \forall\, x,y\in\mathbb{R}^n.$$

实际上，利用 f 的次微分概念，上述不等式可以推广到任意凸函数 $f:\mathbb{R}^n\to(-\infty,\infty]$:

$$\begin{cases}\langle x_1-x_2,y_1-y_2\rangle\geq 0,\\ \forall\, y_1\in\partial f(x_1),\ \forall\, y_2\in\partial f(x_2).\end{cases}\tag{4.3.1}$$

这是因为 $y_1\in\partial f(x_1)$, $y_2\in\partial f(x_2)$ 意味着

$$f(x_2)\geq f(x_1)+\langle x_2-x_1,y_1\rangle,$$

$$f(x_1)\geq f(x_2)+\langle x_1-x_2,y_2\rangle,$$

把这两个不等式相加即得式 (4.3.1)

一般地，对于多值映射 $A:\mathbb{R}^n\to 2^{\mathbb{R}^n}$, 称 A 为单调映射，是指它满足

$$\langle x_1-x_2,y_1-y_2\rangle\geq 0,\quad \forall\, y_1\in Ax_1,\ \forall\, y_2\in Ax_2.$$

A 的图像 $G(A)$ 是指乘积空间 $\mathbb{R}^n \times \mathbb{R}^n$ 中的集合

$$G(A) = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n \mid x \in \operatorname{dom} A,\ y \in Ax\}.$$

于是偶对 $(x,y) \in G(A)$ 指 $x \in \operatorname{dom} A$, 并且 $y \in Ax$. 对于两个多值映射 $A, B : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, $A \subset B$ 是指 $G(A) \subset G(B)$, 即

$$\operatorname{dom} A \subset \operatorname{dom} B,\ \text{并且}\ Ax \subset Bx,\ \forall\, x \in \operatorname{dom} A.$$

称单调映射 A 为极大单调的，是指不存在另一个单调映射 B 真包含 A. 单调映射 $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ 称为闭的，是指其图像 $G(A)$ 是 $\mathbb{R}^n \times \mathbb{R}^n$ 中的闭集；换句话说，如果 $y_k \in Ax_k$, $x_k \to x$, $y_k \to y\,(k \to \infty)$, 则必有 $x \in \operatorname{dom} A$, 并且 $y \in Ax$.

这样我们看到， $\mathbb{R}^n$ 上凸函数 f 的次微分映射 ∂f 是 $\mathbb{R}^n$ 上的一个多值单调映射. 为了进一步讨论次微分映射，我们先对一般的单调映射作一些说明.

引理 4.3.1 设 $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ 为单调映射，那么为了 A 是极大单调的，必须且只须对于某个 $\lambda > 0$ (从而对于一切 $\lambda > 0$) 有

$$R(A + \lambda I) = \mathbb{R}^n, \tag{4.3.2}$$

其中 $R(T)$ 表示映射 T 的值域，而 I 表示 $\mathbb{R}^n$ 中的恒等映射.

证明 首先设式 (4.3.2) 成立，但 A 不是极大单调的，则存在 $(x_0, y_0) \notin G(A)$ 使得

$$\langle x - x_0, y - y_0 \rangle \ge 0, \quad \forall\, (x,y) \in G(A). \tag{4.3.3}$$

但由假设 (4.3.2), $R(A + \lambda I) = \mathbb{R}^n$, 故必有 $x_1 \in \operatorname{dom} A$ 使得

$$y_0+\lambda x_0\in Ax_1+\lambda x_1,$$

即 $y_0+\lambda(x_0-x_1)\in Ax_1$. 在 (4.3.3) 中取 $(x,y)=(x_1,y_0+\lambda(x_0-x_1))$, 得到

$$-\|x_0-x_1\|^2=\langle x_1-x_0,x_0-x_1\rangle\geq 0.$$

从而 $x_1=x_0$, $y_0\in Ax_0$, 这与 $(x_0,y_0)\notin G(A)$ 相抵触. 故 A 是极大单调映射.

现在设 A 是极大单调映射. 今证式 (4.3.2) 成立. 不失一般性, 我们假定 $\lambda=1$. 任意取 $y\in \mathbb{R}^n$, 我们要证明存在 $x\in \operatorname{dom} A$, 使得 $y\in Ax+x$, 即 $y-x\in Ax$. 利用 A 的极大单调性, 为此只须证明存在 $x\in \mathbb{R}^n$, 使得

$$\langle v-(y-x),u-x\rangle\geq 0,\quad \forall\,(u,v)\in G(A).\qquad(4.3.4)$$

对于任意固定的 $(u,v)\in G(A)$, 令

$$M(u,v)=\{x\in \mathbb{R}^n\mid \langle v+x,u-x\rangle\geq\langle y,u-x\rangle\}.$$

今证 $M(u,v)$ 是 $\mathbb{R}^n$ 中的非空有界闭凸集. 事实上, 显然 $u\in M(u,v)$, 从而 $M(u,v)\neq\emptyset$; $M(u,v)$ 的闭性是内积 $\langle\cdot,\cdot\rangle$ 的连续性的直接结果; 而由 $M(u,v)$ 的定义推出

$$\|x\|^2-\langle y,x\rangle+\langle v,x\rangle-\langle u,x\rangle\leq\langle u,v\rangle,\quad \forall\,x\in M(u,v).$$

由此可见 $M(u,v)$ 的有界性. 于是为了式 (4.3.4) 成立, 我们只要证明

$$\cap\{M(u,v)\mid (u,v)\in G(A)\}\neq\emptyset.$$

为此, 依据紧集的有穷覆盖定理, 只须证明对于任意有穷个 $(u_i,v_i)\in G(A), i=1,\cdots,m$, 有

$$\bigcup_{i=1}^{m} M(u_i, v_i) \neq \emptyset.$$

定义 $I\!R^n$ 中有界闭凸集 K,

$$K = \left\{ \lambda = (\lambda_1, \cdots, \lambda_m)^\tau \mid \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1 \right\},$$

以及函数 $f : K \times K \to I\!R$,

$$f(\lambda, \mu) = \sum_{i=1}^{m} \mu_i \langle x(\lambda) + v_i - y, x(\lambda) - u_i \rangle, \quad \forall \lambda, \mu \in K,$$

其中 $x(\lambda) = \sum_{i=1}^{m} \lambda_i u_i$. 不难验证函数 f 关于第一个变元是连续凸的，而关于第二个变元是线性的. 因此根据 Mini-Max 定理 5.2.7 (此定理的证明独立于这里的单调性), 存在 $\lambda^0 \in K$, 使得

$$f(\lambda^0, \mu) \leq \max_{\lambda \in K} f(\lambda, \lambda), \quad \forall \mu \in K.$$

但是

$$\begin{aligned} f(\lambda, \lambda) &= \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j \langle v_i, u_j - u_i \rangle \\ &= \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j \langle v_i - v_j, u_j - u_i \rangle \leq 0, \end{aligned}$$

于是

$$\sum_{i=1}^{m} \mu_i \langle x(\lambda^0) + v_i, x(\lambda^0) - u_i \rangle \leq 0, \quad \forall \mu \in K,$$

即

$$\langle v_i + x(\lambda^0) - y, u_i - x(\lambda^0) \rangle \geq 0, \quad i = 1, \cdots, m.$$

从而 $x(\lambda^0) \in \bigcup_{i=1}^{m} M(u_i, v_i)$. ∎

引理 4.3.2 设 $f: \mathbb{R}^n \to (-\infty, +\infty]$ 为闭的真凸函数，并且满足强制性条件：

$$\lim_{\|x\| \to \infty} f(x) = +\infty. \tag{4.3.5}$$

那么 f 在 $\mathbb{R}^n$ 上达到其最小值，即存在 $x_0 \in \operatorname{dom} f$ 使得

$$f(x_0) \le f(x), \quad \forall x \in \mathbb{R}^n.$$

证明 令

$$\alpha = \inf\{f(x) \mid x \in \mathbb{R}^n\},$$

则 $\alpha < +\infty$, 并且存在 $x_k \in \operatorname{dom} f$ 使得 $\lim_{k\to\infty} f(x_k) = \alpha$. 由假设 (4.3.5), $\{x_k \mid k \ge 1\}$ 是 $\mathbb{R}^n$ 中的有界点列. 于是根据 Weierstrass 紧性定理，不妨设 $\lim_{k\to\infty} x_k = x_0$ (否则可取子列). 再由 f 在 x_0 处的下半连续性，即得

$$f(x_0) \ge \alpha = \lim_{k\to\infty} f(x_k) \ge f(x_0),$$

即

$$f(x_0) = \inf\{f(x) \mid x \in \mathbb{R}^n\}.$$ ∎

定理 4.3.3 设 $f: \mathbb{R}^n \to (-\infty, +\infty]$ 为闭的真凸函数, 那么 f 的次微分映射 ∂f 是 $\mathbb{R}^n$ 中的极大单调映射, 并且

$$R(\partial f + \lambda I) = R^n, \quad \forall \lambda > 0. \tag{4.3.6}$$

证明 根据引理 4.3.1, 只须证明式 (4.3.6). 为简单起见, 不妨设 $\lambda = 1$. 于是对于任意 $y_0 \in \mathbb{R}^n$, 我们要证明方程

$$x + \partial f(x) \ni y_0$$

至少有一个解 $x_0 \in \operatorname{dom}(\partial f)$. 今定义新的闭真凸函数 $\varphi: \mathbb{R}^n \to (-\infty, +\infty]$ 如下：

$$\varphi(x) = \|x\|^2/2 + f(x) - \langle x, y_0 \rangle.$$

注意到 $\mathrm{dom}\,(\partial f) \neq \emptyset$, 可知 f 下有界于某个仿射函数，故

$$\lim_{\|x\| \to \infty} \varphi(x) = +\infty.$$

因此由引理 4.3.2, 存在 $x_0 \in \mathrm{dom}\, f$ 使得

$$\varphi(x_0) \leq \varphi(x), \quad \forall\, x \in I\!R^n.$$

又由定理 4.2.5, 上式等价于 $0 \in \partial\varphi(x_0)$. 为了计算 $\partial\varphi$, 注意由例 4.1.1 和 4.1.7, 有

$$\partial g_1(x) = \{x\}, \quad \partial g_2(x) = \{y_0\},$$

其中 $g_1(x) = \|x\|^2/2$, $g_2(x) = \langle x, y_0 \rangle$. 这样， $0 \in \partial\varphi(x_0)$ 正是说 $y_0 \in \partial f(x_0) + x$. ∎

凸函数 f 的次微分映射 ∂f 实际上有比单调性更强的性质：对于 $(x_i, y_i) \in G(\partial f)$, $i = 0, 1, \cdots, m$, m 为任意自然数，并且记 $(x_{m+1}, y_{m+1}) = (x_0, y_0)$, 则

$$f(x_i) \geq f(x_{i-1}) + \langle x_i - x_{i-1}, y_{i-1} \rangle, \quad i = 1, \cdots, m+1.$$

把这 $m+1$ 个不等式相加得到

$$\left\{ \begin{aligned} &\langle x_1 - x_0, y_0 \rangle + \langle x_2 - x_1, y_1 \rangle \\ &\quad + \cdots + \langle x_{m+1} - x_m, y_m \rangle \leq 0. \end{aligned} \right. \tag{4.3.7}$$

一般地，我们称多值映射 $A: I\!R^n \to 2^{I\!R^n}$ 为循环单调的，是指对于任意 $(x_i, y_i) \in G(A), i = 0, 1, \cdots, m$, m 为任意自然数，并且 $(x_{m+1}, y_{m+1}) = (x_0, y_0)$, 有不等式 (4.3.7) 成立. 这样，凸函数 f 的次微分映射是循环单调的. 注意从定义看出，循环单调映射一定是单调的.

多值映射 A 称为极大循环单调的，是指其图像不能真包含在另一个循环单调映射的图像之中.

现在我们指出，循环单调映射与凸函数的次微分映射密切相关.

定理 4.3.4 设 $A: \mathbb{R}^n \to 2^{\mathbb{R}^n}$ 为多值映射. 那么为了存在一个闭的真凸函数 $f: \mathbb{R}^n \to (-\infty, +\infty]$ 使得 $A \subset \partial f$，必须且只须 A 是循环单调的.

证明 必要性是显然的，因为我们已指出次微分映射 ∂f 的循环单调性. 另一方面，假定 A 是循环单调的，任意取一固定的 $(x_0, y_0) \in G(A)$，并定义 $\mathbb{R}^n$ 上的函数 f，

$$f(x) = \sup\Big\{\langle x - x_m, y_m\rangle + \sum_{i=1}^{m}\langle x_i - x_{i-1}, y_{i-1}\rangle \;\Big|\; (x_i, y_i) \in G(A), 1 \le i \le m\Big\},\ x \in \mathbb{R}^n,$$

这里 m 也是任意的. 由于 f 是某一族仿射函数的上包络，故 f 是闭凸函数. 又 A 的循环单调性意味着 $f(x_0) = 0$，故 f 还是真的. 现在任意取 $(x, y) \in G(A)$，我们证明 $(x, y) \in G(\partial f)$，即 $x \in \mathrm{dom}\,(\partial f)$，并且 $y \in \partial f(x)$. 为此只须证明，对于任意 $\alpha < f(x)$ 和任意 $z \in \mathbb{R}^n$ 有

$$f(z) > \alpha + \langle z - x, y\rangle.$$

事实上，给定 $\alpha < f(x)$，由 f 的定义，存在有穷个 $(x_i, y_i) \in G(A)$，$i = 0, 1, \cdots, m$，使得

$$\alpha < \langle x - x_m, y_m\rangle + \langle x_m - x_{m-1}, y_{m-1}\rangle + \cdots + \langle x_1 - x_0, y_0\rangle.$$

令 $x_{m+1} = x, y_{m+1} = y$，则

$$f(z) \ge \langle z - x_{m+1}, y_{m+1}\rangle + \langle x_{m+1} - x_m, y_m\rangle + \cdots + \langle x_1 - x_0, y_0\rangle > \alpha + \langle z - x, y\rangle, \quad \forall z \in \mathbb{R}^n.$$

由此让 $\alpha \uparrow f(x)$, 得到

$$f(z) \geq f(x) + \langle z - x, y \rangle, \quad \forall z \in \mathbb{R}^n,$$

即 $y \in \partial f(x)$. 这样我们证明了 $A \subset \partial f$. ▌

定理 4.3.5 设 $f: \mathbb{R}^n \to (-\infty, +\infty]$ 为闭的真凸函数，那么 f 的次微分映射 ∂f 是 $\mathbb{R}^n$ 中的极大循环单调映射.

证明 这是定理 4.3.3 和 4.3.4 的直接结果. ▌

我们知道，如果 f, g 是 $\mathbb{R}^n$ 上的两个可微函数，当 $\nabla f(x) = \nabla g(x), \forall x \in \mathbb{R}^n$ 时，f 和 g 只差一加法常数. 人们自然要问，当 f 和 g 是 $\mathbb{R}^n$ 上的两个闭的真凸函数，并且满足 $\partial f = \partial g$ 时，f 和 g 是否也只差一加法常数呢？答案是肯定的. 为了着手讨论，我们设法把它归结到单值次微分映射来处理. 为此先作一些技术性的准备.

设 $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ 为极大单调映射，于是 $R(A + I) = \mathbb{R}^n$. 从而对任意 $x \in \mathbb{R}^n$ 和任意 $\lambda > 0$, 方程

$$u + \lambda A u \ni x$$

至少有一个解 $u \in \operatorname{dom} A$. 实际上，这个解还是唯一的，因为否则，设 $u_1, u_2 \in \operatorname{dom} A$ 为两个解. 于是 $x - u_1 \in Au_1, x - u_2 \in Au_2$. 由 A 的单调性得到

$$0 \leq \langle x - u_1 - (x - u_2), u_1 - u_2 \rangle = -\|u_1 - u_2\|^2,$$

故 $u_1 = u_2$. 这个唯一解记作 $J_\lambda x$. 于是 $J_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ 是单值映射，并且 $\operatorname{dom} J_\lambda = \mathbb{R}^n$. 显然，

$$J_\lambda = (I + \lambda A)^{-1}, \quad \lambda > 0.$$

J_λ 叫做 A 的豫解式. 我们还定义与 A 有关的另一个映射 A_λ:

$$A_\lambda = \frac{1}{\lambda}(I - J_\lambda). \tag{4.3.8}$$

A_λ 称为 A 的 Yosida 近似.

引理 4.3.6 设 $A: \mathbb{R}^n \to 2^{\mathbb{R}^n}$ 为极大单调映射, J_λ 和 A_λ 分别为 A 的豫解式和 Yosida 近似. 那么

(1) J_λ 是 $\mathbb{R}^n$ 的 Lipschitz 映射, 满足

$$\|J_\lambda x - J_\lambda y\| \le \|x - y\|, \quad \forall\, x, y \in \mathbb{R}^n, \forall\, \lambda > 0, \tag{4.3.9}$$

并且

$$\lim_{\lambda \to 0} J_\lambda x = x, \quad \forall\, x \in \mathrm{cl}\,[\mathrm{conv}\,(\mathrm{dom}\, A)]. \tag{4.3.10}$$

(2) $A_\lambda x \in AJ_\lambda x, \quad \forall\, x \in \mathbb{R}^n$.

(3) A_λ 是单调的 Lipschitz 映射, 满足

$$\begin{cases} \|A_\lambda x - A_\lambda y\| \le \frac{1}{\lambda}\|x - y\|, \\ \forall\, x, y \in \mathbb{R}^n, \forall\, \lambda > 0. \end{cases} \tag{4.3.11}$$

证明 (1) 从 J_λ 的定义可知 $\frac{x - J_\lambda x}{\lambda} \in AJ_\lambda x$, $\frac{y - J_\lambda y}{\lambda} \in AJ_\lambda y$. 于是由 A 的单调性得到

$$\langle x - y, J_\lambda x - J_\lambda y\rangle - \|J_\lambda x - J_\lambda y\|^2 \ge 0.$$

由此即得式 (4.3.9). 为证式 (4.3.10), 由 A 的单调性得知,

$$\langle u - J_\lambda x, v - A_\lambda x\rangle \ge 0, \quad \forall\, (u, v) \in G(A).$$

于是对于 $\lambda > 0$,

$$\|x - J_\lambda x\|^2 \le \lambda\langle u - J_\lambda x, v\rangle + \langle x - u, x - J_\lambda x\rangle. \tag{4.3.12}$$

由此可见, 对任意固定的 $x \in \mathbb{R}^n$, 当 $\lambda \to 0$ 时 $\|x - J_\lambda x\|$ 是

有界的. 因此存在序列 $\lambda_k \downarrow 0$ 使得 $x - J_{\lambda_k}x \to y\,(k \to \infty)$. 在式 (4.3.12) 中让 $\lambda = \lambda_k$, 并令 $k \to \infty$ 取极限得到

$$\|y\|^2 \le \langle x-u, y\rangle, \quad \forall\, u \in \operatorname{dom} A.$$

上式显然可以开拓到 $u \in \operatorname{cl}[\operatorname{conv}(\operatorname{dom} A)]$ 也成立. 这样, 特别若 $x \in \operatorname{cl}[\operatorname{conv}(\operatorname{dom} A)]$, 则可取 $u = x$, 得到 $y = 0$. 上面实际上我们证明了从 $x - J_\lambda x$ 出发的任意序列 $\{x - J_{\lambda_k}x\}$ 中都可以找到一个子序列 $\{x - J_{\lambda_k'}x\}$ 收敛于 0. 从而式 (4.3.12) 成立.

(2) $A_\lambda x \in A(J_\lambda x)$ 实际上在 (1) 中已经证明了. 最后证明 (3), 由 A 的单调性, 有

$$\begin{aligned}\langle A_\lambda x - A_\lambda y, x - y\rangle &= \lambda\langle A_\lambda x - A_\lambda y, A_\lambda x - A_\lambda y\rangle \\ &\quad + \langle A_\lambda x - A_\lambda y, J_\lambda x - J_\lambda y\rangle \\ &\ge \lambda\|A_\lambda x - A_\lambda y\|^2 \ge 0, \quad \forall\, x, y \in \mathbb{R}^n,\end{aligned}$$

即得 A_λ 的单调性. 同时从上式也推出 (4.3.11) 成立. ∎

设 $f: \mathbb{R}^n \to (-\infty, +\infty]$ 为闭真凸函数, 于是 $A = \partial f$ 是 $\mathbb{R}^n$ 中的极大单调映射, $R(A+I) = \mathbb{R}^n$. 对于每一个 $\lambda > 0$, 定义函数 $f_\lambda: \mathbb{R}^n \to (-\infty, +\infty]$ 如下:

$$f_\lambda(x) = \inf\left\{\|u-x\|^2/2\lambda + f(u) \mid u \in \mathbb{R}^n\right\}, \quad \forall\, x \in \mathbb{R}^n. \tag{4.3.13}$$

显然, f_λ 是 $\mathbb{R}^n$ 上处处有穷的闭的真凸函数. 仿照引理 4.3.2 的证明, 可知 f_λ 在 $\mathbb{R}^n$ 上达到最小值. 对任意固定的 $x \in \mathbb{R}^n$ 和 $\lambda > 0$, 令

$$g_{x,\lambda}(u) = \|u-x\|^2/2\lambda + f(u),$$

则

$$\partial g_{x,\lambda}(u) = \frac{u-x}{\lambda} + \partial g(u), \quad \forall\, u \in \mathbb{R}^n.$$

利用定理 4.2.5, $g_{x,\lambda}(u)$ 作为 u 的函数在 $\mathbb{R}^n$ 上达到最小值的 u 满足 $0 \in \frac{u-y}{\lambda} + \partial f(u)$, 即 $u = J_\lambda x$. 因此,

$$f_\lambda(x) = \|J_\lambda x - x\|^2/2\lambda + f(J_\lambda x), \quad \forall\, x \in \mathbb{R}^n, \forall\, \lambda > 0.$$

定理 4.3.7 设 $f : \mathbb{R}^n \to (-\infty, +\infty]$ 为闭的真凸函数, 记 $A = \partial f$. 那么由式 (4.3.13) 定义的函数 f_λ 在 $\mathbb{R}^n$ 上可微, 并且 $A_\lambda(x) = \partial f_\lambda(x), \forall\, x \in \mathbb{R}^n$. 此外, 有

$$\begin{cases} f_\lambda(x) = \|J_\lambda x - x\|^2/2\lambda + f(J_\lambda x), \\ \forall\, x \in \mathbb{R}^n, \lambda > 0, \end{cases} \tag{4.3.14}$$

$$\lim_{\lambda \to 0} f_\lambda(x) = f(x), \quad \forall\, x \in \mathbb{R}^n, \tag{4.3.15}$$

$$f(J_\lambda x) \le f_\lambda(x) \le f(x), \quad \forall\, \lambda > 0, \forall\, x \in \mathbb{R}^n. \tag{4.3.16}$$

在证明这个定理之前, 先证明一个引理. 我们已经指出, 对于 $\mathbb{R}^n$ 上的真凸函数 f,

$$\operatorname{dom}(\partial f) \subset f.$$

一般来说, $\operatorname{dom}(\partial f)$ 未必是凸集. 但是我们有

引理 4.3.8 设 $f : \mathbb{R}^n \to (-\infty, +\infty]$ 为闭的真凸函数, 那么,

$$\operatorname{cl}[\operatorname{dom}(\partial f)] = \operatorname{cl}[\operatorname{conv}(\operatorname{dom} \partial f)] = \operatorname{cl}(\operatorname{dom} f).$$

证明 由于 $\operatorname{dom} f$ 是凸集, 只须证明

$$\operatorname{dom} f \subset \operatorname{cl}(\operatorname{dom}(\partial f)).$$

记 $A = \partial f$, 任取 $x \in \operatorname{dom} f$, 则 $J_\lambda x \in \operatorname{dom}(\partial f), \forall\, \lambda > 0$, 并且 $A_\lambda x \in \partial f(J_\lambda x)$. 因此,

$$f(x) \ge f(J_\lambda x) + \langle x - J_\lambda x, A_\lambda x \rangle,$$

即

$$\|x - J_\lambda x\|^2 \le \lambda(f(x) - f(J_\lambda x)).$$

注意到 f 下有界于某个仿射函数，并且 $J_\lambda x$ 当 $\lambda \to 0$ 时是有界的，可知 $\lim\limits_{\lambda\to 0} J_\lambda x = x$. 这表明 $x \in \mathrm{cl}(\mathrm{dom}(\partial f))$. ∎

定理 4.3.7 的证明 定理前面的说明已经证明了式 (4.3.14), 而不等式 (4.3.16) 则是 f_λ 的定义和式 (4.3.14) 的结果. 为证明式 (4.3.15), 我们考虑两种情形. 如果 $x \in \mathrm{dom}\, f$, 则由引理 4.3.6 和 4.3.8, $\lim\limits_{\lambda\to 0} J_\lambda x = x$. 从而由式 (4.3.16) 和 f 的下半连续性得到 $\lim\limits_{\lambda\to 0} f_\lambda(x) = f(x)$. 现在设 $x \notin \mathrm{dom}\, f$, 我们要证 $\lim\limits_{\lambda\to 0} f_\lambda(x) = +\infty$. 假设不然，则存在某个常数 $C \in \mathbb{R}$ 和正数列 $\lambda_k \to 0\,(k \to \infty)$, 使得

$$f_{\lambda_k}(x) \le C, \quad \forall\, k \ge 1.$$

由此利用式 (4.3.14) 可得 $\lim\limits_{\lambda\to 0} J_\lambda x = x$, 并且 $f(J_{\lambda_k} x) \le C$. 然后再次利用 f 的下半连续性得到 $f(x) \le C < \infty$, 这与 $x \notin \mathrm{dom}\, f$ 矛盾.

最后证明 f_λ 的可微性. 设 $x, y \in \mathbb{R}^n$, 由于 $A_\lambda x \in \partial f(J_\lambda x)$, 我们有

$$f(J_\lambda y) - f(J_\lambda x) \ge \langle J_\lambda y - J_\lambda x, A_\lambda x\rangle.$$

于是

$$\begin{aligned} f_\lambda(y) - f_\lambda(x) \ge \frac{1}{2}[&\|A_\lambda y\|^2 - \|A_\lambda x\|^2 \\ &+2\langle A_\lambda x - A_\lambda y, A_\lambda x\rangle] + \langle A_\lambda x, y - x\rangle, \end{aligned}$$

即

$$f_\lambda(y) - f_\lambda(x) - \langle A_\lambda x, y - x\rangle \ge \frac{\lambda}{2}\|A_\lambda y - A_\lambda x\|^2 \ge 0.$$

交换 x 和 y 的位置，可得

$$\begin{aligned} & f_\lambda(y) - f_\lambda(x) - \langle A_\lambda x, y - x\rangle \\ \leq & \langle y - x, A_\lambda y - A_\lambda x\rangle \leq \frac{1}{\lambda}\|y - x\|^2, \end{aligned}$$

这表明 f_λ 在 x 处可微，并且 $\nabla f_\lambda(x) = A_\lambda x$，即 $A_\lambda x = \partial f_\lambda(x), \forall x \in \mathbb{R}^n$. ∎

推论 4.3.9 设 $f : \mathbb{R}^n \to (-\infty, +\infty]$ 为闭的真凸函数. 如果 $\partial f = \partial g$, 则存在常数 c, 使得 $f = g + c$.

证明 由定理 4.3.7, $\partial f_\lambda = \partial g_\lambda, \forall \lambda > 0$. 由于 f_λ 和 g_λ 在 $\mathbb{R}^n$ 上可微，存在常数 c_λ, 使得 $f_\lambda = g_\lambda + c_\lambda$. 设 $x_0 \in \operatorname{dom} f = \operatorname{dom} g$, 则 $c_\lambda = f_\lambda(x_0) - g_\lambda(x_0) \to f(x_0) - g(x_0) = c\,(\lambda \to 0)$. 因此 $f(x) = g(x) + c, \forall x \in \mathbb{R}^n$. ∎

在结束本节之前，我们列举一些凸函数及相应的循环单调映射 (即次微分映射).

例 4.3.1 $\mathbb{R}$ 中的单调映射总是循环单调的. 事实上，设 $\varphi : \mathbb{R} \to 2^{\mathbb{R}}$ 为单调映射，并设 $x_0, x_1, \cdots, x_m = x_0 \in \operatorname{dom}\varphi$. 不失一般性，我们可以假定 $x_0 < x_1 < \cdots < x_{m-1}$, 并设 $y_i \in \varphi(x_i), i = 1, \cdots, m$. 由 φ 的单调性可得 $y_m \leq y_1 \leq \cdots \leq y_{m-1}$. 因此，

$$\sum_{i=1}^m (x_i - x_{i-1})y_{i-1} = \sum_{i=1}^m (x_i - x_{i-1})(y_{i-1} - y_{m-1}) \leq 0.$$

现在设 φ 是 $\mathbb{R}$ 中的极大单调映射，于是它是极大循环单调的，故存在一个下半连续真凸函数 $f : \mathbb{R} \to (-\infty, +\infty]$ 使得 $\varphi = \partial f$. 显然存在 a, b, $-\infty \leq a \leq b \leq +\infty$ 使得

$$(a, b) \subset \operatorname{dom}\varphi \subset \operatorname{dom} f \subset [a, b].$$

令

$$\tilde{\varphi}(x) = \inf\{y \mid y \in \varphi(x)\}, \quad \forall\, x \in \operatorname{dom}\varphi,$$

则 $\tilde{\varphi}$ 是 (a,b) 中的增函数，并且 $\varphi(x) = [\tilde{\varphi}(x-0), \tilde{\varphi}(x+0)]$；如果 $a \in \operatorname{dom}\varphi$（相应地 $b \in \operatorname{dom}\varphi$），则 $\varphi(a) = (-\infty, \tilde{\varphi}(x+0)]$（相应地 $\varphi(b) = [\tilde{\varphi}(x-0), +\infty)$）。最后任取一点 $x_0 \in \operatorname{dom}\varphi$，则

$$f(x) = \begin{cases} f(x_0) + \displaystyle\int_0^x \tilde{\varphi}(\xi)d\xi, & \forall\, x \in [a,b], \\ +\infty, & \forall\, x \notin [a,b]. \end{cases}$$

下面是有关 $\varphi, \varphi_\lambda, f, f_\lambda$ 的几个具体例子及相应的图像.

例 4.3.2

$$\varphi(x) = \begin{cases} \mathbb{R}, & x = 0, \\ \emptyset, & x \neq 0; \end{cases} \qquad f(x) = \begin{cases} 0, & x = 0, \\ +\infty, & x \neq 0; \end{cases}$$

$$\varphi_\lambda(x) = x/\lambda, \quad x \in \mathbb{R}, \qquad f_\lambda(x) = x^2/2\lambda, \quad x \in \mathbb{R}.$$

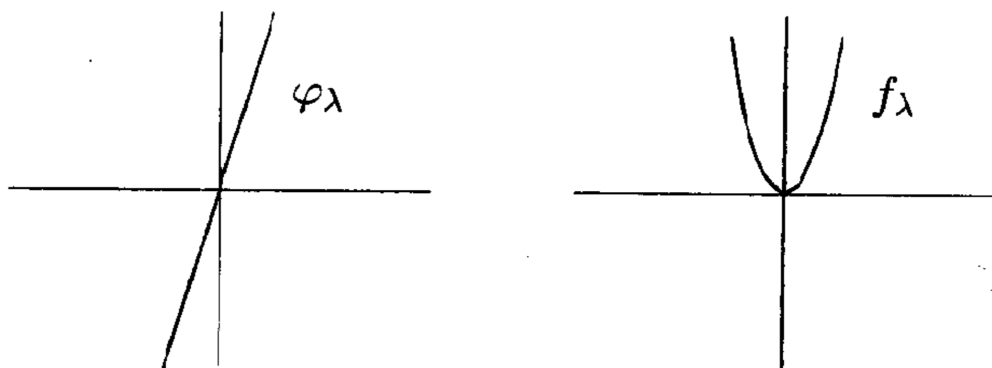


图 4.3.1

例 4.3.3

$$\varphi(x)=\begin{cases}\emptyset, & |x|>1,\\ (-\infty,0], & x=-1,\\ \{0\}, & |x|<1,\\ [0,+\infty), & x=+1,\end{cases}\quad f(x)=\begin{cases}0, & |x|<1,\\ +\infty, & |x|>1;\end{cases}$$

$$\varphi_\lambda(x)=\begin{cases}\frac{x+1}{\lambda}, & x\le -1,\\ 0, & |x|<1,\\ \frac{x-1}{\lambda}, & x\ge 1,\end{cases}\quad f_\lambda(x)=\begin{cases}\frac{(x+1)^2}{2\lambda}, & x\le -1,\\ 0, & |x|<1,\\ \frac{(x-1)^2}{2\lambda}, & x\ge +1.\end{cases}$$

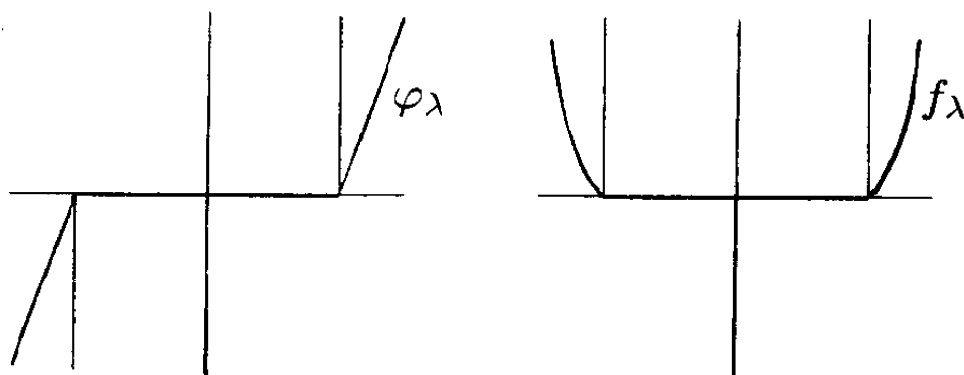


图 4.3.2

例 4.3.4

$$\varphi(x)=kx\ (k\ge 0),\ x\in\mathbb{R},\quad f(x)=kx^2/2,\ x\in\mathbb{R};$$

$$\varphi_\lambda(x)=\frac{kx}{1+\lambda k},\ x\in\mathbb{R},\quad f_\lambda(x)=\frac{kx^2}{2(1+\lambda k)},\ x\in\mathbb{R}.$$

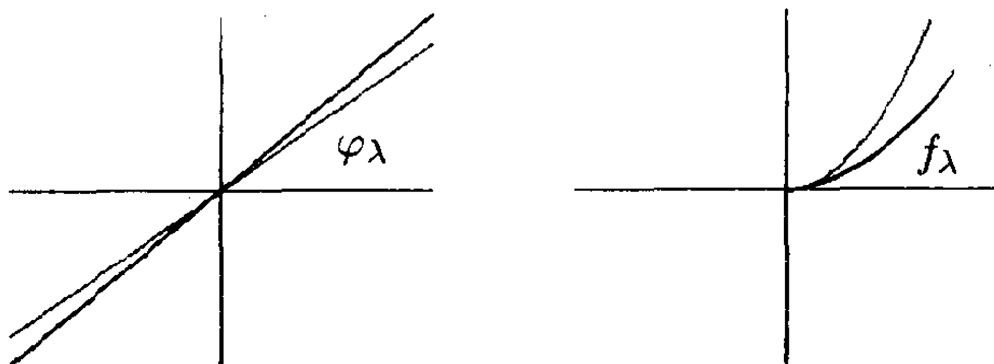


图 4.3.3

例 4.3.5

$$\varphi(x)=\begin{cases}-1, & x<0,\\ [-1,+1], & x=0,\\ +1, & x>0,\end{cases}\qquad f(x)=|x|,\quad x\in I\!R;$$

$$\varphi_\lambda(x)=\begin{cases}-1, & x\le-\lambda,\\ x/\lambda, & |x|<\lambda,\\ +1, & x\ge\lambda,\end{cases}\quad f_\lambda(x)=\begin{cases}-x-\lambda/2, & x\le-\lambda,\\ x^2/2\lambda, & |x|<\lambda,\\ x-\lambda/2, & x\ge\lambda.\end{cases}$$

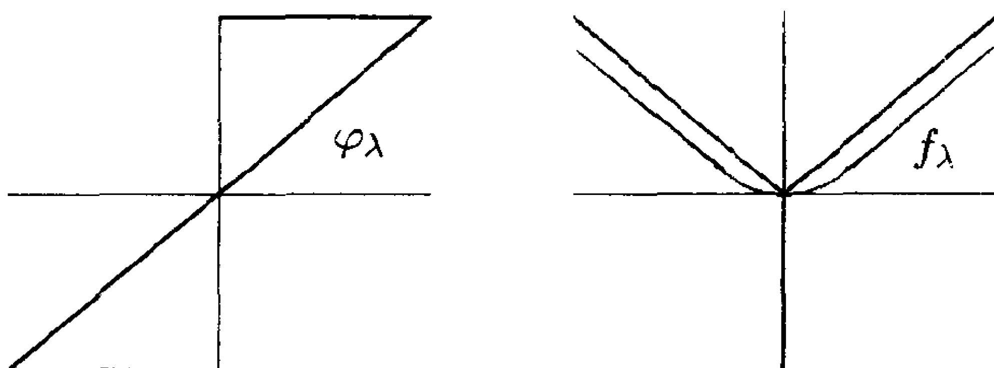


图 4.3.4

在进一步讨论线性循环单调变换之前，我们先证明一个有用的结论.

定理 4.3.10 设 $B: I\!R^n \to I\!R^n$ 为定义在全 $I\!R^n$ 上的一个单值连续的单调映射，那么 B 必是极大单调的.

证明 用反证法. 设 B 不是极大单调的, 于是存在 $x, y \in \mathbb{R}^n, y \neq Bx$, 使得

$$\langle B(u) - y, u - x \rangle \geq 0, \quad \forall u \in \mathbb{R}^n.$$

特别对于任意的 $z \in \mathbb{R}^n$ 和 $\lambda > 0$, 取 $u = x + \lambda z$, 从上式可得

$$\langle B(x + \lambda z) - y, z \rangle \geq 0, \quad \forall z \in \mathbb{R}^n, \ \forall \lambda > 0.$$

让 $\lambda \downarrow 0$, 由 B 的连续性, 从上式得到

$$\langle B(x) - y, z \rangle \geq 0, \quad \forall z \in \mathbb{R}^n.$$

由此可见, $y = B(x)$, 与假设相矛盾. ∎

例 4.3.6 线性循环单调变换. 设 $A : \mathbb{R}^n \to \mathbb{R}^n$ 为线性变换 (即 $n \times n$ 矩阵). 今证 A 为循环单调的充分必要条件是: A 为非负定对称矩阵, 并且这时 $A = \partial f$, 而 f 为与 A 相对应的二次函数

$$f(x) = \langle Ax, x \rangle, \quad \forall x \in \mathbb{R}^n.$$

事实上, 当 A 为非负定对称矩阵时, 由上式定义的 f 是 $\mathbb{R}^n$ 上的连续可微函数, 并且 $A = \partial f$. 现在设 A 循环单调, 于是由定理 4.3.10, A 是极大循环单调的. 因此依定理 4.3.7, 存在 $\mathbb{R}^n$ 上闭的真凸函数 g 使得 $A = \partial g$, 并且不妨取 $g(0) = 0$. 我们知道 A_λ 是 g_λ 的微商. 故

$$\frac{d}{dt} g_\lambda(tx) = \langle A_\lambda(tx), x \rangle = t \langle A_\lambda x, x \rangle,$$

$$\forall x \in \mathbb{R}^n, \forall t \in \mathbb{R}.$$

因此

$$g_\lambda(x) = g_\lambda(x) - g_\lambda(0) = \int_0^1 t\langle A_\lambda x, x\rangle dt = \frac{1}{2}\langle A_\lambda x, x\rangle.$$

由此可见 g_λ 的微商是

$$\partial g_\lambda = \nabla g_\lambda = \frac{1}{2}(A_\lambda + A_\lambda^\tau).$$

但 $\partial g_\lambda = A_\lambda$, 故 $A_\lambda^\tau = A_\lambda$. 又在目前情况下， $A_\lambda = J_\lambda A$, 因此由引理 4.3.6 可知， $A_\lambda x \to Ax\,(\lambda \to 0), \forall x \in \mathbb{R}^n$. 从而 $A = A^\tau$.

例 4.3.7 闭凸集示性函数的次微分. 设 K 是 $\mathbb{R}^n$ 中的非空闭凸集. 在例 4.1.6 中我们曾指出

$$\partial\delta(x|K) = \begin{cases} N_K(x), & x \in K, \\ \emptyset, & x \notin K, \end{cases}$$

其中 N_K 为 K 在 x 处的法向锥. 为了找出 $\partial\delta(\cdot|K)$ 的豫解式和 Yosida 近似，我们先定义闭凸集 K 上的投影映射.

容易看出，对于任意闭凸集 $M \subset \mathbb{R}^n$ 和任意向量 $x \in \mathbb{R}^n$, 存在唯一的向量 $y \in M$ 使得

$$\|y - x\| = \inf\{\|y - z\| \mid z \in M\}. \qquad (4.3.17)$$

事实上，令

$$f_x(z) = \|x - z\|, \quad \forall\, z \in \mathbb{R}^n,$$

则 $f_x(z)$ 作为 z 的函数是 $\mathbb{R}^n$ 的连续凸函数，并且满足强制性条件：

$$\lim_{\|z\|\to\infty} f_x(z) = +\infty.$$

因此根据引理 4.3.2, 存在 $y \in M$ 使得式 (4.3.17) 成立. 至

于 y 的唯一性则从 f_x 的严格凸性得出. 记 $y = P_M x$, 于是我们得到一个新的映射 $P_M : \mathbb{R}^n \to M$. P_M 叫做 $\mathbb{R}^n$ 到 M 上的投影映射.

注意对于闭凸集 $K \subset \mathbb{R}^n$, $\lambda > 0$ 和 $x \in \mathbb{R}^n$, 如果令

$$g_{\lambda,x}(z) = \frac{1}{2}\|z - x\|^2 + \lambda\delta(z|K), \quad \forall\, z \in \mathbb{R}^n,$$

则

$$y = P_K x \Longleftrightarrow g_{\lambda,x}(y) = \inf\{g_{\lambda,x}(z) \mid z \in \mathbb{R}^n\}.$$

但依照定理 4.2.5, 又有

$$g_{\lambda,x}(y) = \inf\{g_{\lambda,x}(z) \mid z \in \mathbb{R}^n\} \Longleftrightarrow 0 \in \partial g_{\lambda,x}(y).$$

又从例 4.1.1 我们知道,

$$\partial g_{\lambda,x}(y) = y - x + \lambda\partial\delta(y|K).$$

因此

$$\begin{aligned} y = P_K x &\Longleftrightarrow x \in y + \lambda\partial\delta(y|K) \\ &\Longleftrightarrow y = (I + \lambda\partial\delta(\cdot|K))^{-1}(x). \end{aligned}$$

这样我们证明了

$$P_K = (I + \lambda\delta(\cdot|K))^{-1}, \quad \forall\, \lambda > 0.$$

于是 $\delta(\cdot|K)$ 的 Yosida 近似 $(\partial(\cdot|K))_\lambda(x) = \frac{1}{\lambda}(x - P_K x)$ 是函数

$$(\delta(\cdot|K))_\lambda(x) = \frac{1}{2\lambda}\|x - P_K x\|^2$$

的导数.

第五章 凸分析的应用

凸分析所需要的预备知识相对来说是比较少的，因此比较容易被实际应用部门的数学工作者接受. 凸分析的应用领域是十分广泛的，如数学规划 (包括凸规划和线性规划), 对策论，数理经济学，最优控制理论，等等. 这些领域本身实际上早已形成了各自独立的学科分支. 我们在本章中仅仅简要地介绍凸函数极值问题、凸规划和线性规划问题. 作为某种基础， 5.1 节中先介绍 Helly 定理和不等式组解的问题. 至于控制理论方面的应用，由于要涉及较多的别的学科的知识，这里就不作介绍了，但它所要用到的凸分析的一些知识则基本上由本书所包含了.

5.1 Helly 定理和不等式组

设 $\mathcal{F}$ 是一族集合，m 为一固定的自然数. 如果 $\mathcal{F}$ 中任意 m 个集合具有某种性质 P, 则整个集合族 $\mathcal{F}$ 就具有性质 Q. 这一类结果通称为 Helly 型定理. 最早也是最广为人知的这一类结果之一是 Edward Helly(1884–1943) 在 1913 年发现的. 此后出现了 Helly 定理的各种各样的证明和推广. 这些证明方法中有些是几何的，有些是代数的，另有一些则使用了对偶性. 我们首先采用 Radon 的办法，从他的一个基本结果出发来证明 Helly 定理.

定理 5.1.1 (**Radon定理**) 设 $M = \{x_1, x_2, \cdots, x_m\}$

是 $\mathbb{R}^n$ 中任一有限集. 如果 $m \geq n+2$, 则 M 可以分解成两个不相交的子集 M_1 和 M_2, 使得 $\mathrm{conv}\, M_1 \cap \mathrm{conv}\, M_2 \neq \emptyset$.

证明 由于 $m \geq n+2$, 依定理 1.1.5, M 是仿射相关的, 即存在不全为 0 的实数 $\lambda_1, \cdots, \lambda_m$ 使得

$$\sum_{i=1}^{m} \lambda_i x_i = 0, \quad \sum_{i=1}^{m} \lambda_i = 0.$$

诸 λ_i 中至少有两个有相反的正负号, 于是不妨设对于某个 k, $1 \leq k < m$,

$$\lambda_1 \geq 0, \cdots, \lambda_k \geq 0, \lambda_{k+1} < 0, \cdots, \lambda_m < 0.$$

这样, 我们有

$$\lambda = \lambda_1 + \cdots + \lambda_k = -(\lambda_{k+1} + \cdots + \lambda_m) > 0.$$

令

$$x = \sum_{i=1}^{k} (\lambda_i/\lambda) x_i = \sum_{i=k+1}^{m} -(\lambda_i/\lambda) x_i,$$

则 x 是 $x_1, \cdots, x_k$ 的凸组合, 即 $x \in \mathrm{conv}\, \{x_1, \cdots, x_k\}$. 类似地我们也有 $x \in \mathrm{conv}\, \{x_{k+1}, \cdots, x_m\}$. 因此, 如果设

$$M_1 = \{x_1, \cdots, x_k\}, \quad M_2 = \{x_{k+1}, \cdots, x_m\},$$

则 $\mathrm{conv}\, M_1 \cap \mathrm{conv}\, M_2 \neq \emptyset$. ∎

定理 5.1.2 (Helly定理) 设 $\mathcal{F} = \{B_1, \cdots, B_m\}$ 是 $\mathbb{R}^n$ 中由 m 个凸集组成的集族, 假定 $m \geq n+1$. 如果 $\mathcal{F}$ 的任意 $n+1$ 个集合都有非空交, 那么整个集族 $\mathcal{F}$ 也有非空交, 即 $\bigcap_{i=1}^{m} B_i \neq \emptyset$.

证明 采用关于集族 $\mathcal{F}$ 中凸集数 m 的归纳法. 当 $m = n+1$ 时定理当然成立. 现在假定, 对于由 $m = r-1 (\geq n+1)$

个凸集组成的任意集族定理成立. 考察由 r 个凸集组成的集族 $\mathcal{F}=\{B_1,\cdots,B_r\}$, 并且设 $\mathcal{F}$ 中任意 $n+1$ 个集有非空交. 于是依据归纳假设, $\cap\{B_i \mid 1\le i\le r, i\ne j\}\ne\emptyset$, 从而存在 $x_j\in\cap\{B_i \mid 1\le i\le r, i\ne j\}$, $\forall j=1,\cdots,r$. 由于 $r\ge n+2$, Radon 定理适用于集合 $M=\{x_1,\cdots,x_r\}$, 得到 M 的两个不相交的子集 M_1 和 M_2, 使得

$$\operatorname{conv} M_1\cap\operatorname{conv} M_2\ne\emptyset.$$

不失一般性, 我们可以设

$$M_1=\{x_1,\cdots,x_k\},\quad M_2=\{x_{k+1},\cdots,x_r\},$$

其中 $1\le k<r$. 任取点 $x\in\operatorname{conv} M_1\cap\operatorname{conv} M_2$, 今证 $x\in B_i$, $i=1,\cdots,r$. 事实上, 对于 $i\le k$, 我们有 $x_i\in B_{k+1}\cap\cdots\cap B_r$. 因此, 由于每个 B_i 是凸集, 我们有

$$x\in\operatorname{conv}\{x_1,\cdots,x_k\}\subset B_{k+1}\cap\cdots\cap B_r.$$

同样地, 对于 $i>k$, 有 $x_i\in B_1\cap\cdots\cap B_k$. 因此

$$x\in\operatorname{conv}\{x_{k+1},\cdots,x_r\}\subset B_1\cap\cdots\cap B_k.$$ ∎

Helly 定理可用于有穷个凸函数不等式组的解的存在性的证明.

推论 5.1.3 设 $f_1,\cdots,f_m\,(m\ge n+1)$ 为 $\mathbb{R}^n$ 上的凸函数, $M\subset\mathbb{R}^n$ 为给定的非空凸集. 考虑如下凸函数不等式组:

$$f_1(x)<0,\cdots,f_k(x)<0,f_{k+1}(x)\le 0,\cdots,f_m(x)\le 0.$$

假定上述不等式组中任意由不超过 $n+1$ 个不等式构成的子组在 M 中都有解, 那么整个不等式组在 M 中也有解.

证明 令

$$
\begin{aligned}
M_0 &= M, \\
M_i &= \{x \in \mathbb{R}^n \mid f_i(x) < 0\}, i = 1, \cdots, k, \\
M_i &= \{x \in \mathbb{R}^n \mid f_i(x) \leq 0\}, i = k+1, \cdots, m,
\end{aligned}
$$

然后应用定理 5.1.2 于集族 $\{M_i \mid i = 0, 1, \cdots, m\}$ 即得所需要的结论. ∎

许多作者对上述 Helly 定理作了大量的推广和应用，这些推广和应用与凸函数不等式组密切相关. Helly 定理中出现的集合必须是凸集，这容易举出反例来说明. 同时，"Helly 数"$n+1$ 也不能减少，例如 $\mathbb{R}^2$ 中三角形任意两条边有公共点，但整个三条边却没有公共点. Helly 定理可以推广到无穷多个集合组成的集族，但必须附加条件.

今后我们称两个命题 P 和 Q 二择一成立是指命题 P 和 Q 有且只有一个成立.

定理 5.1.4 设 $f_1, \cdots, f_m$ 为 $\mathbb{R}^n$ 上的真凸函数，M 为 $\mathbb{R}^n$ 的非空凸集. 假定 $\operatorname{ri} M \subset \operatorname{dom} f_i, 1 = 1, \cdots, m$. 那么下列两个命题二择一成立：

(1) 存在某个 $x \in M$, 使得

$$f_1(x) < 0, \cdots, f_m(x) < 0;$$

(2) 存在不全为 0 的非负实数 $\lambda_1, \cdots, \lambda_m$, 使得

$$\lambda_1 f_1(x) + \cdots + \lambda_m f_m(x) \geq 0, \quad \forall x \in M.$$

证明 先假定 (1) 成立，并取满足 (1) 的任意 $x \in M$, 以及任意非负实数 $\lambda_1, \cdots, \lambda_m$. 于是表达式

$$\lambda_1 f_1(x) + \cdots + \lambda_m f_m(x)$$

的每一项是非正的，其中对应于非零 λ_i 的项是负的. 因此当 λ_i 不全为 0 时 整个表达式必为负值. 这说明 (2) 不能成立.

现在假定 (1) 不成立. 定义映射 $T: \mathbb{R}^n \to \mathbb{R}^m$,

$$T(x) = (f_1(x), \cdots, f_m(x))^\tau, \quad \forall x \in \mathbb{R}^n,$$

以及 $\mathbb{R}^n$ 中的两个集合 A 和 B,

$$A = \{z \in \mathbb{R}^m \mid \exists x \in M \text{ 使得 } T(x) < z\},$$

$$B = \{z \in \mathbb{R}^m \mid z \leq 0\},$$

这里对于两个向量 $y = (\eta_1, \cdots, \eta_m)^\tau$ 和 $z = (\zeta_1, \cdots, \zeta_m)^\tau$, $y < z$ 是指 $\eta_i < \zeta_i, 1 \leq i \leq m$, 而 $y \leq z$ 是指 $\eta_i \leq \zeta_i, 1 \leq i \leq m$. A 显然是 $\mathbb{R}^m$ 中的非空凸集，并且由于 (1) 不成立， A 与 B 必不相交，即 $A \cap B = \emptyset$. 于是 A 和 B 可以用 $\mathbb{R}^m$ 中的某个超平面真分离，即存在非零向量 $z^0 = (\lambda_1, \cdots, \lambda_m)^\tau$ 和某个实数 α, 使得

$$\begin{cases} \alpha \leq \langle z^0, z \rangle = \lambda_1\zeta_1 + \cdots + \lambda_m\zeta_m, \\ \forall z = (\zeta_1, \cdots, \zeta_m)^\tau \in A, \end{cases} \tag{5.1.1}$$

$$\begin{cases} \alpha \geq \langle z^0, z \rangle = \lambda_1\zeta_1 + \cdots + \lambda_m\zeta_m, \\ \forall z = (\zeta_1, \cdots, \zeta_m)^\tau \in B. \end{cases} \tag{5.1.2}$$

由于 B 是非正象限，从式 (5.1.2) 可知 $\alpha \geq 0, \lambda_i \geq 0, i = 1, \cdots, m$. 于是从式 (5.1.1) 又可知，只要 $x \in M$, 使得 $f_i(x) < \zeta_i, i = 1, \cdots, m$, 就有

$$\lambda_1\zeta_1 + \cdots + \lambda_m\zeta_m \geq 0.$$

因此, 对于任意 $x \in D = M \cap \operatorname{dom} f_1 \cap \cdots \cap \operatorname{dom} f_m$, 我们有

$$\lambda_1(f_1(x)+\varepsilon)+\cdots+\lambda_m(f_m(x)+\varepsilon)\geq 0,\quad \forall\varepsilon>0.$$

由于 $\varepsilon>0$ 的任意性，从上式可得

$$\lambda_1 f_1(x)+\cdots+\lambda_m f_m(x)\geq 0,\quad \forall x\in D.$$

这样，凸函数 $f=\lambda_1 f_1+\cdots+\lambda_m f_m$ 在凸集 D 上非负且处处有穷. 因此根据推论 2.4.6, $f(x)\geq 0,\forall x\in \mathrm{cl}\,D$. 但由假设 $\mathrm{ri}\,M\subset \mathrm{dom}\,f_i$ 可知 $\mathrm{ri}\,M\subset D$, 于是

$$M\subset \mathrm{cl}\,(\mathrm{ri}\,M)\subset \mathrm{cl}\,D.$$

因此 (2) 成立. ∎

下一个定理涉及到仿射函数的特殊性质.

定理 5.1.5 设 $f_1,\cdots,f_k$ 为 $I\!R^n$ 上的真凸函数，$f_{k+1},\cdots,f_m$ 为 $I\!R^n$ 上的仿射函数，而 M 为 $I\!R^n$ 上的非空凸集. 假定不等式组

$$f_{k+1}(x)\leq 0,\cdots,f_m(x)\leq 0$$

至少有一个解 $x\in \mathrm{ri}\,M$. 那么下列两个命题二择一成立:

(1) 存在某个 $x\in M$, 使得

$$f_1(x)<0,\cdots,f_k(x)<0,f_{k+1}(x)\leq 0,\cdots,f_m(x)\leq 0;$$

(2) 存在不全为 0 的非负数 $\lambda_1,\cdots,\lambda_k$, 以及非负数 $\lambda_{k+1},\cdots,\lambda_m$, 使得

$$\lambda_1 f_1(x)+\cdots+\lambda_m f_m(x)\geq 0,\quad \forall x\in M.$$

证明 同定理 5.1.4 的证明一样，显然当 (1) 成立时 (2) 不能成立. 现在假定 (1) 不成立，我们要证明 (2) 成立. 为此令 (式中 $z=(\zeta_1,\cdots,\zeta_m)^\tau$)

$$A = \{z \in \mathbb{R}^m \mid \exists x \in M \text{ 使得 } f_i(x) < \zeta_i, 1 \le i \le k;$$
$$f_i(x) = \zeta_i, k+1 \le i \le m\},$$
$$B = \{z \in \mathbb{R}^m \mid z \le 0\}.$$

从假设可知 A 是非空凸集，并且由于 (1) 不成立， A 和 B 不相交. 从而 A 和 B 可以用一个超平面真分离，即存在实数 α 和非零向量 $z^0 = (\lambda_1, \cdots, \lambda_m)^\tau$, 使得

$$\begin{cases} \alpha \le \langle z^0, z\rangle = \lambda_1\zeta_1 + \cdots + \lambda_m\zeta_m, \\ \forall\, z = (\zeta_1, \cdots, \zeta_m)^\tau \in A, \end{cases} \tag{5.1.3}$$

$$\begin{cases} \alpha \ge \langle z^0, z\rangle = \lambda_1\zeta_1 + \cdots + \lambda_m\zeta_m, \\ \forall\, z = (\zeta_1, \cdots, \zeta_m)^\tau \in B, \end{cases} \tag{5.1.4}$$

并且式 (5.1.3) 至少对于一个 $z \in A$ 成立严格不等号. 从 5.1.4 推出 $\alpha \ge 0, \lambda_i \ge 0, \forall\, 1 \le i \le m$. 式 (5.1.3) 等价于对任意 $x \in M, \zeta_i > f_i(x), i = 1, \cdots, k$, 成立

$$\lambda_1\zeta_1 + \cdots + \lambda_k\zeta_k + \lambda_{k+1}f_{k+1}(x) + \cdots + \lambda_m f_m(x) \ge \alpha.$$

于是若令 $D = M \cap \operatorname{dom} f_1 \cap \cdots \cap \operatorname{dom} f_k$, 则 $\forall\, x \in D$,

$$\lambda_1 f_1(x) + \cdots + \lambda_k f_k(x) + \lambda_{k+1}f_{k+1}(x) + \cdots + \lambda_m f_m(x) \ge \alpha.$$

由于凸函数 $f = \lambda_1 f_1 + \cdots + \lambda_m f_m$ 满足 $f(x) \ge \alpha, \forall\, x \in D$, 推论 2.4.6 推出 $f(x) \ge \alpha, \forall\, x \in \operatorname{cl} D$. 但依据假设 $\operatorname{ri} M \subset D$, 所以 $f(x) \ge \alpha, \forall\, x \in M$. 从而

$$\lambda_1 f_1(x) + \cdots + \lambda_m f_m(x) \ge 0, \quad \forall\, x \in M.$$

为了证明 (2) 成立，剩下只要证明 $\lambda_1, \cdots, \lambda_k$ 不全为 0. 假

定 $\lambda_1 = \cdots = \lambda_k = 0$. 于是 $f = \lambda_{k+1} f_{k+1} + \cdots + \lambda_m f_m$, 从而 f 是仿射函数. 根据定理假设, 至少有一点 $x \in \mathrm{ri}\, M$ 使得 $f_i(x) \le 0, \forall k+1 \le i \le m$. 对于这样的 x 我们有 $f(x) \le 0$. 但是 $f(x) \ge \alpha \ge 0, \forall x \in M$, 因此 $\alpha = 0$, 并且 f 在 M 上的下确界在 $\mathrm{ri}\, M$ 上达到. 这样, f 作为仿射函数, 在 M 上只能取常值, 即 $f(x) = 0, \forall x \in M$. 另一方面, 我们知道至少有一点 $z = (\zeta_1, \cdots, \zeta)^{\tau} \in A$, 使得

$$\alpha < \lambda_1 \zeta_1 + \cdots + \lambda_m \zeta_m.$$

于是存在一点 $x \in M$ 使得

$$0 \le \alpha < \lambda_{k+1} f_{k+1}(x) + \cdots + \lambda_m f_m(x) = f(x),$$

这与 $f(x) = 0, \forall x \in M$ 相矛盾. ∎

下一个重要定理涉及到凸函数不等式的解的存在性.

定理 5.1.6 设 $\{f_i \mid i \in I\}$ 为 $I\!R^n$ 上的一族闭的真凸函数, 其中 I 为任一指标集. 设 $M \subset I\!R^n$ 为任一非空闭凸集. 假定诸 f_i 的任何公共回收方向都不是 M 的回收方向. 那么如下两个命题二择一成立:

(1) 存在 $x \in M$, 使得

$$f_i(x) \le 0, \quad \forall i \in I;$$

(2) 存在非负数 $\lambda_i \,(i \in I)$, 其中仅有穷个非零, 使得对某个 $\varepsilon > 0$ 有

$$\sum_{i \in I} \lambda_i f_i(x) \ge \varepsilon, \quad \forall x \in M.$$

此外在 (2) 成立的情况下, 实际上可选择乘子 λ_i 使得其中至多 $n+1$ 个为非零.

证明 根据定理 2.4.19, 把每一个函数 f_i 加上示性函数 $\delta(\cdot|M)$, 我们可以把定理化成 $M = I\!R^n$ 的情形. 显然 (1)

和 (2) 不可能同时成立. 现在假定 (1) 不成立, 我们要证明 (2) 成立.

设 g 是由

$$h = \operatorname{conv}\{f_i^* \mid i \in I\}$$

所生成的正齐性凸函数. 依照定理 3.2.10, g^* 是凸集 $\{x \in \mathbb{R}^n \mid h^*(x) \leq 0\}$ 的示性函数. 由于每一个 f_i 是闭的, 根据定理 3.4.8, 有

$$h^* = \sup\{f_i^{**} \mid i \in I\} = \sup\{f_i \mid i \in I\}.$$

因此 g^* 是集合

$$D = \{x \in \mathbb{R}^n \mid f_i(x) \leq 0, \forall\, i \in I\}$$

的示性函数. 但是由于 (1) 不成立, $D = \emptyset$, 所以 g^* 是常值函数 $+\infty$, 并且 $\operatorname{cl} g = g^{**}$ 必是常值函数 $-\infty$. 特别 $(\operatorname{cl} g)(0) = -\infty$.

首先, 如果 $g(0) = -\infty$, 则 $h(0) < 0$. 根据定理 2.3.5 及其后的注, 可知存在非负数 λ_i, 其中至多 $n+1$ 个为非零, 使得

$$\sum_{i \in I} \lambda_i z_i = 0, \quad \sum_{i \in I} \lambda_i f_i^*(z_i) < 0.$$

为了记号简单起见, 我们把这些非零的 λ_i 重新排序记成 $\lambda_1, \cdots, \lambda_m$ $(m \leq n+1)$, 并令 $y_i = \lambda_i z_i$. 于是

$$y_1 + \cdots + y_m = 0,$$

$$(f_1^* \lambda_1)(y_1) + \cdots + (f_m^* \lambda_m)(y_m) < 0.$$

从而根据下端卷积的定义,

$$(f_1^* \lambda_1 \square \cdots \square f_m^* \lambda_m)(0) < 0.$$

令 $f=\lambda_1 f_1+\cdots+\lambda_m f_m$, 依据定理 3.4.7,

$$\begin{aligned} f^* &= \mathrm{cl}\left((\lambda_1 f_1)^*\Box\cdots\Box(\lambda_m f_m)^*\right) \\ &= \mathrm{cl}\left(f_1^*\lambda_1\Box\cdots\Box f_m^*\lambda_m\right). \end{aligned}$$

因此 $f^*(0)\le(f_1^*\lambda_1\Box\cdots\Box f_m^*\lambda_m)(0)<0$. 但依据定义,

$$f^*(0)=\sup\{\langle x,0\rangle-f(x)\mid x\in I\!R^n\}=-\inf f.$$

由此可见 $\inf f>0$, 即存在某个 $\varepsilon>0$, 使得

$$\lambda_1 f_1(x)+\cdots+f_m(x)\ge\varepsilon,\quad \forall\, x\in I\!R^n,$$

这表明，这一组 λ_i 使得 (2) 成立.

现在我们证明 $g(0)=(\mathrm{cl}\,g)(0)$, 从而确实有 $g(0)=-\infty$. 事实上, g 的有效定义域 $\mathrm{dom}\,g$ 是由诸集 $\mathrm{dom}\,f_i^*, i\in I$ 所生成的凸锥. 如果 $0\in\mathrm{ri}\,(\mathrm{dom}\,g)$, 则从定理 2.5.1 知 $g(0)=(\mathrm{cl}\,g)(0)$. 如果 $0\notin\mathrm{ri}\,(\mathrm{dom}\,g)$, 则依据定理 1.5.5, 我们可以用一个超平面把 0 和 $\mathrm{dom}\,g$ 分离，即存在非零向量 y, 使得

$$\langle y,z\rangle\le 0,\quad \forall\, z\in\mathrm{dom}\,g.$$

于是

$$\langle y,z\rangle\le 0,\quad \forall\, z\in\mathrm{dom}\,f_i^*,\ \forall\, i\in I,$$

这表明 $(f_i0^+)(y)\le 0,\forall i\in I$, 即 y 是每个 f_i 的回收方向 (见定理 3.2.11). 这与定理的假设相矛盾. 因此 $0\in\mathrm{ri}\,(\mathrm{dom}\,g)$. ▮

应用定理 5.1.6, 无穷凸不等式组的存在性问题可以化成有穷凸不等式组的存在性问题.

推论 5.1.7 设 $\{f_i\mid i\in I\}$ 为 $I\!R^n$ 上的一族闭的真凸函数，其中 I 为任一指标集. 设 $M\subset I\!R^n$ 为任一非空闭凸集. 假定诸 f_i 的公共回收方向都不是 M 的回收方向,

并且对于任意的 $\varepsilon > 0$ 和 I 中任意 m ($\leq n+1$) 个标号 $i_1, \cdots, i_m$, 不等式组

$$f_{i_1}(x) < \varepsilon, \cdots, f_{i_m}(x) < \varepsilon \tag{5.1.5}$$

至少有一个解 $x \in M$. 那么存在一个 $x \in M$ 满足

$$f_i(x) \leq 0, \quad \forall i \in I.$$

证明 只要证明在目前的假设下，定理 5.1.6 中的 (2) 不成立. 事实上，如果 (2) 成立，则存在 I 中 m ($\leq n+1$) 个标号 $i_1, \cdots, i_m$ 和相应的正数 λ_{i_k}, 使得对于某个 $\delta > 0$ 有

$$\sum_{k=1}^{m} \lambda_{i_k} f_{i_k}(x) \geq \delta, \quad \forall x \in M.$$

令 $\lambda = \lambda_{i_1} + \cdots + \lambda_{i_m}, \varepsilon = \delta/\lambda$. 那么

$$\sum_{k=1}^{m} (\lambda_{i_k}/\lambda)(f_{i_k}(x) - \varepsilon) \geq 0 \quad \forall x \in M.$$

这是不可能的，因为在目前的假设下，存在一个 $x \in M$ 使得 (5.1.5) 成立. ▮

利用定理 5.1.6 还可以把 Helly 定理推广到无穷集族的情形. 当然这时对集族中的集合要作相应的限制.

定理 5.1.8 (Helly定理) 设 $\mathcal{F} = \{M_i \mid i \in I\}$ 是 $\mathbb{R}^n$ 中一族非空闭凸集，其中 I 为任一指标集. 假定诸集 M_i 没有任何公共的回收方向. 如果 $\mathcal{F}$ 的不超过 $n+1$ 个集组成的任意子族都有非空交，则整个集族 $\mathcal{F}$ 也有非空交.

证明 定义函数 $f_i = \delta(\cdot|M_i)$, 并令 $M = \mathbb{R}^n$. 于是，诸集 M_i 没有公共回收方向等价于诸函数 f_i 没有公共回收

方向，而关于 $\mathcal{F}$ 的不超过 $n+1$ 个集组成的子族都有非空交的假设使得定理 5.1.6 的相应假设成立. 这样从定理 5.1.6 知存在 $x \in I\!R^n$, 使得

$$\delta(x \mid M_i) \leq 0, \quad \forall i \in I,$$

这意味着 $x \in M_i, \forall i \in I$. ▮

特别当集族 $\mathcal{F}$ 中至少有一个集合有界时，整个集族显然没有公共回收方向.

习 题 5.1

5.1.1 试在 $I\!R^1$ 和 $I\!R^2$ 中举例说明 Helly 定理中集合的凸性的假设是必不可少的.

5.1.2 设 $\mathcal{F} = \{B_i \mid i \in I\}$ 为 $I\!R^n$ 中的一族紧凸集，并且 $\mathcal{F}$ 至少含有 $n+1$ 个集合. 假定 $I\!R^n$ 的紧凸集 K 具有性质：对 $\mathcal{F}$ 中由任意 $n+1$ 个集组成的子族，都存在 K 的一个平移使之包含在这子族的所有 $n+1$ 个集合中. 试证：存在 K 的一个平移使之包含在 $\mathcal{F}$ 的所有集合中. （提示：对于每一个 B_i, 设 $A_i = \{x \in I\!R^n \mid x + K \subset B_i\}$, 那么 A_i 是紧凸集.）

5.1.3 设 $\mathcal{F}$ 为 $I\!R^n$ 中的一族紧凸集，并且 $\mathcal{F}$ 至少含有 $n+1$ 个集合. 假定 $I\!R^n$ 的紧凸集 K 具有性质：对 $\mathcal{F}$ 中由任意 $n+1$ 个集组成的子族，都存在 K 的一个平移使之包含这子族的所有 $n+1$ 个集合. 试证存在 K 的一个平移使之包含 $\mathcal{F}$ 的所有集合.

5.1.4 设 $\mathcal{F}$ 为 $I\!R^n$ 中的一族紧凸集，并且 $\mathcal{F}$ 至少含有 $n+1$ 个集合. 假定对 $\mathcal{F}$ 中由任意 $n+1$ 个集组成的子族，$I\!R^n$ 中存在一点使它到该子族的每一个集合的距离不超过某个给定的正数 ε. 试证：$I\!R^n$ 中存在一点使它到 $\mathcal{F}$ 每一个集合的距离不超过 ε.

5.1.5 设 $\{x, y, z, w\}$ 为 $I\!R^n$ 的子集，满足

$$\operatorname{conv}\{x, y, z\} \cap \operatorname{conv}\{x, y, w\} \cap \operatorname{conv}\{x, z, w\} = \{x\}.$$

试证：$x \in \operatorname{conv}\{y, z, w\}$.

5.1.6 试证如下 Radon 定理的逆命题：设 $M=\{x_1,\cdots,x_m\}$ 为 $\mathbb{R}^n$ 中的有限子集，$2\le m\le n+1$. 如果 M 是仿射无关的，则当 M 分解为两个不相交的子集 M_1 和 M_2 时都有

$$\operatorname{conv} M_1\cap\operatorname{conv} M_2=\emptyset.$$

5.1.7 设 $M\subset\mathbb{R}^n$ 为非空凸集，$f_1,\cdots,f_m:M\to\mathbb{R}$ 为 m 个真凸函数. 如果存在 $\varepsilon>0$, 使得

$$M=\bigcup_{i=1}^{m}\{x\in M\mid f_i(x)>\varepsilon\},$$

试证：必存在 m 个非负实数 $\lambda_1,\cdots,\lambda_m$, $\lambda_1+\cdots+\lambda_m=1$, 使得

$$\sum_{i=1}^{m}\lambda_i f_i(x)\ge\varepsilon,\forall\, x\in M.$$

5.2 凸函数的极小值问题

本节中我们研究凸集 $M\subset\mathbb{R}^n$ 上凸函数 g 的极小值问题. 不失一般性，我们可以假定 g 是真凸函数. 显然在 M 上求 g 的极小等价于在整个 $\mathbb{R}^n$ 上求凸函数

$$f(x)=g(x)+\delta(x|M)$$

的极小. 我们先讨论 $\mathbb{R}^n$ 上真凸函数 f 的无约束极小值问题，然后再考察 $f=g+\delta(\cdot|M)$ 这种特殊形式的凸函数. f 在整个 $\mathbb{R}^n$ 上的下确界记作

$$\inf f=\inf\{f(x)\mid x\in\mathbb{R}^n\}.$$

注意对于 $\mathbb{R}^n$ 上的凸函数 f 来说，f 在某点达到局部极小等价于 f 在该点达到整体极小. 事实上，如果 z 是 f 的局部极小值点，即存在某个 $\varepsilon>0$, 使得

$$f(z)\le f(x),\quad\forall\, x\in B(z,\varepsilon),$$

则必有

$$f(z) \leq f(x), \quad \forall x \in I\!R^n.$$

这是因为，如果 z 不是 f 的整体极小值点，则必有某个 $y \in \operatorname{dom} f$, 使得 $f(y) < f(z)$. 于是由 f 的凸性，

$$f(\lambda y+(1-\lambda)z) \leq \lambda f(y)+(1-\lambda)f(z) < f(z), \quad \forall 0 < \lambda < 1$$

但是对于充分小的 $\lambda > 0$, $\lambda y + (1-\lambda)z \in B(z,\varepsilon)$, 这与 z 是 f 的局部极小值点相矛盾.

对于 $I\!R^n$ 上真凸函数 f, 我们知道水平集

$$S_\alpha(f) = \{x \in I\!R^n \mid f(x) \leq \alpha\}$$

是 $I\!R^n$ 中的凸子集，并且若 f 还是闭的，则 $S_\alpha(f)$ 也是闭的. 显然 $\operatorname{dom} f = \bigcup_{\alpha \in I\!R^n} S_\alpha(f)$. 在 $I\!R^n$ 上求 f 的极小值相当于在 $\operatorname{dom} f$ 上求 f 的极小值. 下确界 $\inf f$ 的特征是: $\alpha < \inf f \Longrightarrow S_\alpha(f) = \emptyset$. 当 $\alpha = \inf f$ 时， $S_\alpha(f)$ 恰好由所有达到 f 的下确界的点 x 组成，这个水平集叫做 f 的极小集. 显然，研究 f 的极小集的性质具有重要意义. 真凸函数 f 的极小集总是 $I\!R^n$ 的一个凸子集 (可能是空集), 并且若 f 还是闭的，则它也是闭的. 此外，若 f 在其有效定义域上是严格凸的，则其极小集至多由一点组成.

从次梯度的定义可知，为了点 x 属于真凸函数 f 的极小集，必须且只须 $0 \in \partial f(x)$, 即 $x^* = 0$ 是 f 在 x 处的一个次梯度. 因此对于具体的极小值问题来说， 4.2 节中计算次梯度的一些公式是非常有用的.

首先，我们把以上讨论的有关凸函数极小的一些结果综合成如下一个定理:

定理 5.2.1 设 $f : I\!R^n \to (-\infty, +\infty]$ 是一个闭的真凸函数. 那么，

(1) $\inf f = -f^*(0)$. 于是 f 下有界 $\Longleftrightarrow 0 \in \operatorname{dom} f^*$.

(2) f 的极小集是 $\partial f^*(0)$. f 达到其下确界当且仅当 f^* 在 0 处次可微. 特别当 $0 \in \operatorname{ri}(\operatorname{dom} f^*)$ 时 f^* 在 0 处次可微.

(3) 为了 f 的下确界是有穷值但达不到, 必须且只须 $f^*(0)$ 有穷, 并且存在一个 $y \in I\!R^n$, 使得 $f_+^{*\prime}(0; y) = -\infty$.

(4) 为了 f 的极小集是一个非空有界集, 必须且只须 $0 \in \operatorname{int}(\operatorname{dom} f^*)$, 该条件等价于 f 没有任何回收方向.

(5) 为了 f 的极小集由唯一向量 x 组成, 必须且只须 f^* 在 0 点可微, 并且 $x = \nabla f^*(0)$.

(6) 诸水平集 $S_\alpha(f)$ 中的非空集均有相同的回收锥, 并且就是 f 的回收锥, 它是由 $\operatorname{dom} f^*$ 生成的凸锥的极化锥.

证明 (1) 是 $f^*(0)$ 的定义.

(2) 从 $\partial f^*(0)$ 的定义和定理 4.1.3 推出.

(3) 注意, $\partial f^*(0) = \emptyset$ 当且仅当 $\partial f^*(0)$ 的承托函数为常值函数 $-\infty$. 依据定理 4.1.3, 这个承托函数等于 $\operatorname{cl}(f_+^{*\prime}(0; \cdot))$. 但为了凸函数 $f_+^{*\prime}(0; \cdot)$ 的闭包恒等于 $-\infty$, 必须且只须此函数在某点 y 取值 $-\infty$, 即 $f_+^{*\prime}(0; y) = -\infty$.

(4) 从 (2) 、定理 4.1.3 和定理 2.3.9 推出.

(5) 是定理 4.1.4 的结果.

(6) 的第一个结论从定理 2.4.18 推出. 为了证明 (6) 的第二个结论, 任意取 $\alpha > \inf f$, 我们有

$$\begin{aligned} S_\alpha(f) &= \{x \mid f(x) \leq \alpha\} \\ &= \{x \mid \langle x, x^* \rangle - f^*(x^*) \leq \alpha, \forall\, x^* \in \operatorname{dom} f*\}. \end{aligned}$$

于是为了 $y \in 0^+(S_\alpha(f))$, 即

$$x + \lambda y \in S_\alpha(f), \quad \forall\, x \in S_\alpha(f), \forall\, \lambda \geq 0,$$

必须且只须

$$\langle y, x^* \rangle \leq 0, \quad \forall\, x^* \in \dim f^*.$$

而这正是说 y 属于由 $\mathrm{dom}\, f^*$ 所生成的凸锥的极化锥. ∎

定理 5.2.2 设 $f : \mathbb{R}^n \to (-\infty, +\infty]$ 为闭的真凸函数, 并且没有回收方向. 那么 f 的下确界是有穷的并且被达到. 此外, 对于任意 $\varepsilon > 0$, 存在 $\delta > 0$ 使得集合 $\{x \mid f(x) \leq \inf f + \delta\}$ 与 f 的极小集之间的距离不超过 ε, 即 f 的极小集中至少有一个向量 z 使得 $\|z - x\| < \varepsilon, \forall\, x \in S_\alpha(f)$, 这里 $\alpha = \inf f + \delta$. 这时 f 的极小集是一个非空闭有界凸集.

证明 从定理 5.2.1 之 (4) 可知, $\inf f$ 有穷并且被达到. 由于 f 没有回收方向, 从定理 5.2.1 之 (6) 可知其所有水平集 $S_\alpha(f)$ 是有界的. 设 M 为 f 的极小集. 任取 $\varepsilon > 0$, 则 $M_\varepsilon = M + \varepsilon B(0,1)$ 为开集. 令

$$U_\alpha = M_\varepsilon^c \cap S_\alpha(f), \quad \alpha = \inf f + \delta, \delta > 0,$$

这里 M_ε^c 为 M_ε 在 $\mathbb{R}^n$ 中的补集. 如果每一个 U_α 都是非空的, 则诸集 $U_\alpha\, (\delta > 0)$ 形成 $\mathbb{R}^n$ 中的一个有界闭子集的集套, 从而诸 U_α 至少有一个公共点 x

$$\begin{cases} f(x) \leq \inf f + \delta, & \forall\, \delta > 0, \\ x \notin M + \varepsilon B(0,1). \end{cases}$$

这是不可能的, 因为上述第一个关系表明 $x \in M$. 这样, 至少有一个 $\delta > 0$ 使得 $U_\alpha = \emptyset$. 对于这个 δ, 水平集 $\{x \in \mathbb{R}^n \mid f(x) \leq \inf f + \delta\}$ 完全包含在 M_ε 中. ∎

推论 5.2.3 设 f 为 $\mathbb{R}^n$ 上没有任何回收方向的闭真凸函数. 设 $\mathbb{R}^n$ 中序列 $x_1, x_2, \cdots$ 满足

$$\lim_{k \to \infty} f(x_k) = \inf f. \tag{5.2.1}$$

那么 $x_1, x_2, \cdots$ 是有界序列，并且其所有极限点均属于 f 的极小集.

推论 5.2.4 设 f 为 $I\!R^n$ 上闭真凸函数，假定其下确界在唯一点 x 处达到. 如果序列 $x_1, x_2, \cdots$ 使得 (5.2.1) 成立，则 $\lim\limits_{k\to\infty} x_k = x$.

证明 当 f 的极小集 (也是 f 的一个水平集 $S_\alpha(s), \alpha = \inf f$) 由单点组成时，$f$ 不可能有回收方向. 于是从定理 5.2.2 得知推论的结论成立. ∎

向量序列 $\{x_k\} \subset I\!R^n$ 叫做 f 的极小化序列，是指它满足

$$\lim_{k\to\infty} f(x_k) = \inf f.$$

对于凸函数 f 来说，一个重要的问题是其极小化序列 $\{x_k\}$ 的收敛性，因为如果 $\lim\limits_{k\to\infty} x_k = x_0$, 则当 f 下半连续时，x_0 必属于 f 的极小集，即 $f(x_0) = \inf f$. 推论 5.2.3 和 5.2.4 给出了凸函数 f 的极小化序列收敛的一些充分条件. 我们在引理 4.3.2 也曾指出，对于闭的真凸函数 f, 当 f 满足强制性条件：

$$\lim_{\|x\|\to\infty} f(x) = +\infty,$$

则 f 的极小集不空，并且 f 的极小化序列是有界的，其任意收敛子列的极限必属于 f 的极小集.

定理 5.2.5 设 g 为 $I\!R^n$ 上闭真凸函数，M 为 $I\!R^n$ 的一非空闭凸集. 如果 g 和 M 没有公共的回收方向，则 g 在 M 上达到其下确界.

证明 设 $f(x) = g(x) + \delta(x|M)$, 则 g 在 M 上的下确界等于 f 在 $I\!R^n$ 上的下确界. 如果 f 恒等于 $+\infty$, 则其下确界在整个 M 上达到. 如果 f 不恒等于 $+\infty$, 则 f 为闭的

真凸函数，其回收方向是 g 和 M 的公共回收方向 (见定理 2.4.19). 因此，当 g 和 M 没有公共的回收方向时，f 也没有回收方向. 从而依定理 5.2.2, f 达到其在 $\mathbb{R}^n$ 上的下确界. ∎

在许多情况下，凸函数 f 的极小集中的点可以通过凸函数的次微分理论来刻划. 例如，设 $f_1, f_2, \cdots, f_m$ 是 $\mathbb{R}^n$ 上的真凸函数，$\lambda_1, \cdots, \lambda_m$ 是非负实数. 我们要在整个 $\mathbb{R}^n$ 求凸函数

$$f = \lambda_1 f_1 + \cdots + \lambda_m f_m$$

的极小. 我们知道，f 在点 x 处达到极小的充分必要条件是 $0 \in \partial f(x)$. 在相当一般的条件下 (见定理 4.2.6) 我们有

$$\partial f(x) = \lambda_1 \partial f_1(x) + \cdots + \lambda_m \partial f_m(x), \quad \forall x \in \mathbb{R}^n.$$

这样，上述充分必要条件就化成

$$0 \in \lambda_1 \partial f_1(x) + \cdots + \lambda_m \partial f_m(x).$$

进一步分析 f_k 和 ∂f_k 便可以得到一系列新的结果.

定理 5.2.6 设 g 为 $\mathbb{R}^n$ 上的真凸函数，$M \subset \mathbb{R}^n$ 为非空凸集. 那么为了 g 在 M 上于 x 处达到极小，只须存在一向量 $x^* \in \partial g(x)$, 使得 $-x^*$ 是 M 在 x 处的法向量. 当 $\mathrm{ri}\,(\mathrm{dom}\, g) \cap \mathrm{ri}\, M \neq \emptyset$ 时上述条件也是必要的.

证明 我们只要在整个 $\mathbb{R}^n$ 求函数

$$f = g + \delta(\cdot|M)$$

的极小. 依定理 4.2.1, 为了 f 在 $\mathbb{R}^n$ 于 x 处达到极小，条件

$$0 \in \partial g(x) + \partial \delta(x|M) \tag{5.2.2}$$

总是充分的. 但 $\partial \delta(x|M)$ 恰好是 M 在 x 处的法向锥，因

此式 (5.2.2) 等价于存在一向量 $x^* \in \partial g(x)$ 使得 $-x^*$ 是 M 在 x 处的法向量. 当 $\mathrm{ri}\,(\mathrm{dom}\, g) \cap \mathrm{ri}\, M \neq \emptyset$ 时, 依定理 4.2.6,

$$\partial f(x) = \partial g(x) + \partial \delta(x|M).$$

因此上述条件也是必要的. ∎

当 g 在 x 可微时, $\partial g(x)$ 由单个向量 $\nabla g(x)$ 组成. 定理 5.2.6 中的条件就是: $-\nabla g(x)$ 是 M 在 x 处的法向量. 特别地若 M 是某个子空间, 上述条件就是说: $x \in M$ 并且 $\nabla g(x) \perp M$.

作为定理 5.2.6 的应用, 我们来讨论寻找一凸集 $M \subset \mathbb{R}^n$ 到一给定点 a 的最近点的问题. 实际上这正是在凸集 M 上求可微函数

$$g(x) = \frac{1}{2}\|x - a\|^2$$

的极小问题. 定理 5.2.6 的相交的假设: $\mathrm{ri}\,(\mathrm{dom}\, g) \cap \mathrm{ri}\, M \neq \emptyset$ 在这里显然是满足的. 因此为了点 x 是 M 到 a 的最近点, 必须且只须向量

$$-\nabla g(x) = a - x$$

是 M 在 x 处的法向量, 即

$$\langle z - x, x - a\rangle \geq 0, \quad \forall\, z \in M.$$

在结束本节之前, 我们介绍极小极大定理 (MiniMax 定理), 也称鞍点定理. 这个定理是博奕论 (对策论) 的基础, 在非线性分析和优化理论中也有重要的应用. 深入讨论这方面的问题可参阅有关专著 (例如, 见 [6]).

我们来分析 $\mathbb{R}^2$ 中的函数 $f(x, y) = y^2 - x^2$ 在原点附近的性状. 在 $x = 0$ 的截面上, $(0, 0)$ 是 $f(0, y)$(作为 y 的函数) 的最小值点; 而在 $y = 0$ 的截面上, $(0, 0)$ 是 $f(x, 0)$(作

为 x 的函数) 的最大值点，即

$$f(x,0) \le f(0,0) \le f(0,y), \quad x \in \mathbb{R}, \forall y \in \mathbb{R}.$$

我们称这样的点 $(0,0)$ 为函数 f 的 "鞍点"(见图 5.2.1).

一般地，对于定义在积集 $A \times B$ 上的实值函数 $f(x,y)$, $(x_0,y_0) \in A \times B$ 称为函数 f 的一个鞍点，是指它满足

$$f(x,y_0) \le f(x_0,y_0) \le f(x_0,y), \quad \forall x \in A, \forall y \in B.$$

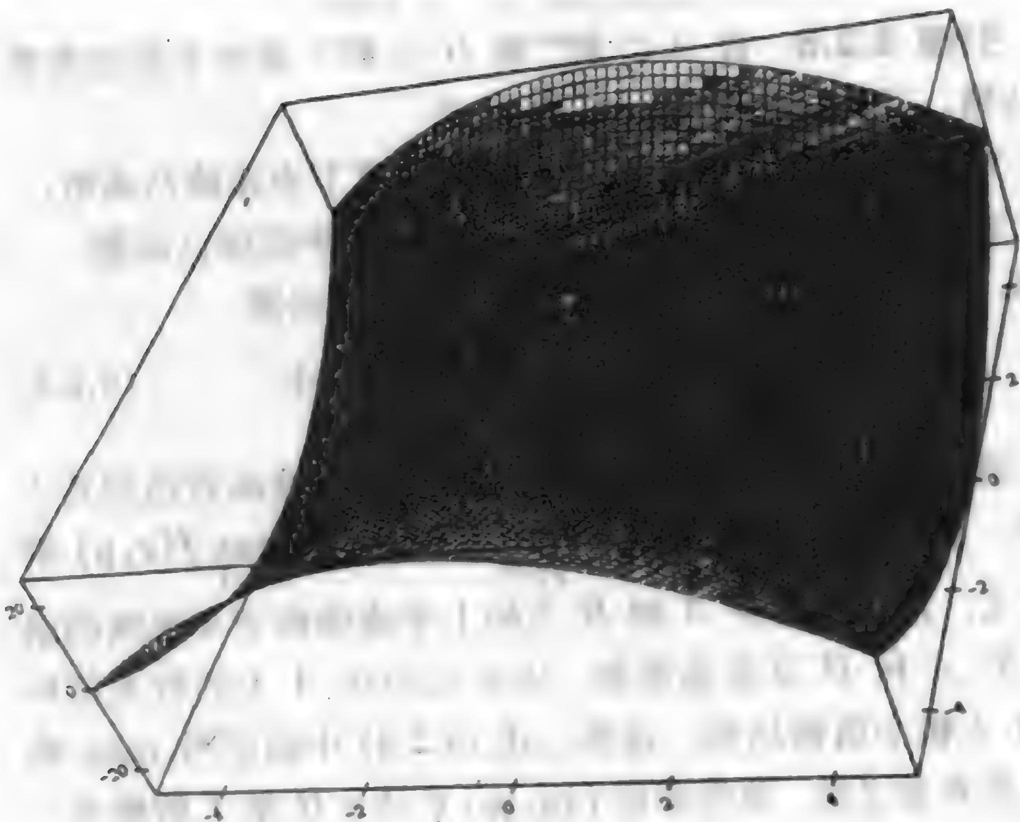


图 5.2.1　函数 $f(x,y) = y^2 - x^2$ 在鞍点 $(0,0)$ 的图形

定理 5.2.7 (极小极大定理)　设 $A \subset \mathbb{R}^n$ 和 $B \subset \mathbb{R}^m$ 是两个非空有界闭凸集，又设 $f: A \times B \to \mathbb{R}$ 满足

$\forall y \in B$, $-f(\cdot,y)$ 是 A 上的有穷值下半连续凸函数，

$\forall x \in A$, $f(x,\cdot)$ 是 B 上的有穷值下半连续凸函数.

那么, f 在 $A\times B$ 上至少有一个鞍点, 即存在一点 $(x_0,y_0)\in A\times B$ 使得

$$f(x,y_0)\le f(x_0,y_0)\le f(x_0,y),\quad \forall x\in A,\forall y\in B, \tag{5.2.3}$$

即

$$\max_{x\in A} f(x,y_0)=f(x_0,y_0)=\min_{y\in B} f(x_0,y).$$

在证明这个定理之前先证明一个引理.

引理 5.2.8 设 $A\subset \mathbb{R}^n$ 和 $B\subset \mathbb{R}^m$ 是两个非空有界闭凸集, 又设 $f: A\times B\to \mathbb{R}$ 满足

$\forall y\in B$, $-f(\cdot,y)$ 是 A 上的有穷值下半连续凸函数,

$\forall x\in A$, $f(x,\cdot)$ 是 B 上的有穷值下半连续凸函数.

那么, 为了 (x_0,y_0) 是 f 的鞍点, 必须且只须

$$\min_{y\in B}\max_{x\in A} f(x,y)=\max_{x\in A}\min_{y\in B} f(x,y). \tag{5.2.4}$$

证明 注意, 定义在紧凸集上的下半连续函数达到其下确界. 从而在引理假设下, $\min\limits_{y\in B} f(x,y)$ 和 $\max\limits_{x\in A} f(x,y)$ 都有意义, 并且分别是 A 和 B 上的上半连续和下半连续凸函数. 但 A 和 B 又都是紧集, 因此它们在 A 上的极大值和在 B 上极小值被达到. 这样, 式 (5.2.4) 中的记号 max 和 min 是有意义的. 首先假定 $(x_0,y_0)\in A\times B$ 是 f 的鞍点, 即式 (5.2.3) 成立. 于是

$$\begin{cases} \min_{y\in B}\max_{x\in A} f(x,y)\le \max_{x\in A} f(x,y_0) \\ \le \min_{y\in B} f(x_0,y)\le \max_{x\in A}\min_{y\in B} f(x,y), \end{cases}$$

即

$$\min_{y\in B}\max_{x\in A} f(x,y)\le \max_{x\in A}\min_{y\in B} f(x,y). \tag{5.2.5}$$

另一方面，从明显成立的不等式

$$\min_{y\in B} f(x,y) \le f(x,y), \quad \forall x \in A, \forall y \in B$$

推出

$$\max_{x\in A}\min_{y\in B} f(x,y) \le \max_{x\in A} f(x,y), \quad \forall y \in B.$$

上式右端再按 y 取极小即得

$$\max_{x\in A}\min_{y\in B} f(x,y) \le \min_{y\in B}\max_{x\in A} f(x,y). \tag{5.2.6}$$

综合式 (5.2.5) 和式 (5.2.6), 得知式 (5.2.4) 成立.

现在设式 (5.2.4) 成立. 今证 f 必有鞍点 $(x_0, y_0) \in A \times B$. 事实上，式 (5.2.4) 成立意味着存在 $x_0 \in A, y_0 \in B$ 使得

$$\min_{y\in B} f(x_0, y) = \max_{x\in A} f(x, y_0).$$

于是有

$$\max_{x\in A} f(x, y_0) = f(x_0, y_0) = \min_{y\in B} f(x_0, y)$$

从而式 (5.2.3) 成立，即 (x_0, y_0) 是 f 的鞍点. ∎

定理 5.2.7 **的证明** 从上述引理 5.2.8 的证明中得知，不等式 (5.2.6) 总是成立的，因此为证鞍点存在，只要再证明不等式 (5.2.5) 成立就够了.

现在假设定理 5.2.7 的条件成立. 记

$$\alpha = \min_{y\in B}\max_{x\in A} f(x,y).$$

对任意固定的 $\varepsilon > 0$ 和 $x \in A$, 令

$$g_x(y) = f(x,y) - \alpha + \varepsilon, \quad \forall y \in B.$$

于是依假设， $g_x(y)$ 作为 y 的函数在 B 上是下半连续的；

并且对任意的 $y \in B$, 存在 $x \in A$ 使得 $g_x(y) > 0$. 这就是说不等式组

$$g_x(y) \leq 0, \quad \forall x \in A,$$

在 B 上无解. 这样，根据定理 5.1.6, 存在一自然数 m, A 中的 m 个元 $x_1, \cdots, x_m$, 以及正数 $\lambda_1, \cdots, \lambda_m$; $\sum_{i=1}^{m} \lambda_i = 1$, 使得

$$g(y) = \sum_{i=1}^{m} \lambda_i g_{x_i}(y) > 0, \quad \forall y \in B.$$

但是

$$g(y) = \sum_{i=1}^{m} \lambda_i f(x_i, y) - \alpha + \varepsilon.$$

于是利用 $-f(x,y)$ 相对于 x 的凸性得到

$$f\Big(\sum_{i-1}^{m} \lambda_i x_i, y\Big) > \alpha - \varepsilon,$$

即

$$\max_{x \in A} \min_{y \in B} f(x,y) \geq \alpha - \varepsilon.$$

但是 $\varepsilon > 0$ 是任意的，所以式 (5.2.5) 成立. ∎

习　题　5.2

5.2.1 设 $f : \mathbb{R}^n \to (-\infty, +\infty]$ 为真凸函数, $M \subset \mathbb{R}^n$ 为非空闭凸集. 假定 $\operatorname{ri}(\operatorname{dom} f) \cap \operatorname{ri} M \neq \emptyset$, 求证: $x_0 \in M$ 是 f 在 M 的极小值点当且仅当存在 $z \in \partial f(x_0)$ 使得

$$\langle z, x - x_0 \rangle \geq 0, \quad \forall x \in M.$$

5.2.2 设 $f : \mathbb{R}^n \to (-\infty, +\infty]$ 为凸函数, $M \subset \mathbb{R}^n$ 为非空闭凸集. 假定 f 在 $x_0 \in M$ 处可微. 求证: 为了 f 在 M 上于 x_0 处达到最小值, 必须且只须

$$\langle \nabla f(x_0), x - x_0 \rangle \geq 0, \quad \forall x \in M.$$

5.2.3 试讨论下列极值问题：

$$f(x) \to \min, \quad Ax \in M,$$

其中 $f: \mathbb{R}^n \to (-\infty, +\infty]$ 为闭真凸函数，$M \subset \mathbb{R}^m$ 为闭凸集，而 $A: \mathbb{R}^n \to \mathbb{R}^m$ 为线性变换.

5.3 凸规划问题

在数学分析中，采用所谓 Lagrange 乘子法可以把 $\mathbb{R}^n$ 上的一个约束极值问题转化成变量较多的另一个无约束极值问题. 这里我们把这一方法应用于凸函数在所谓“凸约束”条件下求极小的问题.

所谓 $\mathbb{R}^n$ 上的凸规划问题 $(\mathcal{P})$ 就是指

$$(\mathcal{P}) \qquad \begin{cases} f_0(x) \to \min, \\ f_i(x) \leq 0, \quad i = 1, \cdots, k, \\ f_j(x) = 0, \quad j = k+1, \cdots, m, \end{cases}$$

其中 $f_0, f_1, \cdots, f_k$ 为 $\mathbb{R}^n$ 上真凸函数，$f_{k+1}, \cdots, f_m$ 为 $\mathbb{R}^n$ 上的仿射函数. 特别当 $k = m$ 时问题 $(\mathcal{P})$ 无等式约束，而当 $k = 0$ 时 $(\mathcal{P})$ 无不等式约束. 因此凸规划问题 $(\mathcal{P})$ 就是在不等式约束 $f_i(x) \leq 0, i = 1, \cdots, k$ 和等式约束 $f_j(x) = 0, j = k+1, \cdots, m$ 之下求目标函数 f_0 的极小的问题. 当然对于凸规划问题 $(\mathcal{P})$ 来说，只有诸函数 f_i 在 $\operatorname{dom} f_0$ 上的值才起作用. 因此不失一般性，在本节后面的

讨论中，我们始终作如下的假设而不特别说明：

$$\begin{cases} \operatorname{ri}(\operatorname{dom} f_j) \supset \operatorname{ri}(\operatorname{dom} f_0), \\ \operatorname{dom} f_j \supset \operatorname{dom} f_0, \quad 1 \le j \le k. \end{cases} \tag{5.3.1}$$

向量 $x \in \mathbb{R}^n$ 叫做 $(\mathcal{P})$ 的可行解，是指 x 满足 $(\mathcal{P})$ 的 m 个约束条件. 于是 $(\mathcal{P})$ 的可行解集 (可能是空集) 是指 $\mathbb{R}^n$ 中的集合

$$M = M_1 \cap \cdots \cap M_m,$$

其中

$$\begin{aligned} M_i &= \{x \in \mathbb{R}^n \mid f_i(x) \le 0\}, \quad i = 1, \cdots, k; \\ M_j &= \{x \in \mathbb{R}^n \mid f_j(x) = 0\}, \quad j = k+1, \cdots, m. \end{aligned}$$

为了今后叙述方便起见，我们总是假定 $M \neq \emptyset$. 于是上述凸规划问题 $(\mathcal{P})$ 可以简写成

$$(\mathcal{P}) \qquad \min f_0(x), \quad x \in M.$$

今定义 $\mathbb{R}^n$ 上新的凸函数 f 如下：

$$f(x) = f_0(x) + \delta(x|M).$$

显然 f 的有效定义域 $\operatorname{dom} f = M \cap \operatorname{dom} f_0$, 并且若 $f_0, \cdots, f_m$ 是闭的，则 f 也是闭的. 注意 f 也可以写成：

$$f(x) = f_0(x) + \delta(x|M_1) + \cdots + \delta(x|M_m). \tag{5.3.2}$$

这样在 M 上求 f_0 的极小等价于在整个 $\mathbb{R}^n$ 上求 f 的极小，即凸规划问题 $(\mathcal{P})$ 也等价于

$$(\mathcal{P}) \qquad f(x) \to \min, \quad x \in \mathbb{R}^n.$$

f 的下确界 $\inf f$ 叫做 $(\mathcal{P})$ 的最优值. 达到 f 的下确界的点

x 叫做 $(\mathcal{P})$ 的最优解. 不难看出, $(\mathcal{P})$ 的最优解集是其可行解集的一个凸子集.

我们先讨论 $k=m$ 和 $\operatorname{dom} f_0 = \mathbb{R}^n$ 的最简单的情况.

定理 5.3.1 设 $f_0, f_1, \cdots, f_m$ 是 $\mathbb{R}^n$ 上的有穷值真凸函数. 假定不等式组

$$f_1(x) < 0, \cdots, f_m(x) < 0 \tag{5.3.3}$$

在 $\mathbb{R}^n$ 上至少有一个解. 考虑如下极值问题:

$$\begin{cases} f_0(x) \to \min, \\ f_i(x) \leq 0, \quad i = 1, \cdots, m. \end{cases} \tag{5.3.4}$$

那么为了式 (5.3.4) 有最优解 $\tilde{x}$, 必须且只须存在实数 $\lambda_1, \cdots, \lambda_m$, 使得如下的所谓 Kuhn-Tucker 条件成立:

(1) $\lambda_i \geq 0,\ f_i(\tilde{x}) \leq 0,\ \lambda_i f_i(\tilde{x}) = 0,\ i = 1, \cdots, m$;

(2) $0 \in \partial f_0(\tilde{x}) + \lambda_1 \partial f_1(\tilde{x}) + \cdots + \lambda_m \partial f_m(\tilde{x})$.

证明 令

$$M_i = \{x \in \mathbb{R}^n \mid f_i(x) \leq 0\}, \quad i = 1, \cdots, m,$$

$$f(x) = f_0(x) + \delta(x|M_1) + \cdots + \delta(x|M_m).$$

我们知道, 为了 $\tilde{x}$ 是凸规划问题 (5.3.4) 的最优解, 必须且只须 $0 \in \partial f(\tilde{x})$. 但式 (5.3.3) 在 $\mathbb{R}^n$ 上有解意味着

$$\operatorname{int} M_1 \cap \cdots \cap \operatorname{int} M_m \neq \emptyset.$$

考虑到 $\operatorname{dom} f_0 = \mathbb{R}^n, \operatorname{dom}(\delta(\cdot|M_i)) = M_i$, 从定理 4.2.6 得到

$$\partial f(x) = \partial f_0(x) + \partial\delta(x|M_1) + \cdots + \partial\delta(x|M_m).$$

但 $\partial\delta(x|M_i)$ 是 M_i 在 x 处的法向锥，并且根据定理 4.2.8,

$$\partial\delta(x|M_i)=\begin{cases}\cup\{\lambda_i\partial f_i(x)\mid\lambda_i\geq 0\}, & 如果\ f_i(x)=0,\\ \{0\}, & 如果\ f_i(x)<0,\\ \emptyset, & 如果\ f_i(x)>0,\end{cases}$$

由此 $\partial f_i(\tilde{x})\neq\emptyset \iff f_i(\tilde{x})\leq 0, i=1,\cdots,m$, 并且为了 $0\in\partial f(\tilde{x})$, 必须且只须存在实数 $\lambda_i, i=1,\cdots,m$ 使得 (1) 和 (2) 成立. ▮

这个定理中的系数 $\lambda_1,\cdots,\lambda_m$ 就是 Lagrange 乘子. 对于一般的凸规划问题 $(\mathcal{P})$, 我们称 $(\lambda_1,\cdots,\lambda_m)^\tau\in\mathbb{R}^m$ 是 $(\mathcal{P})$ 的 Lagrange 乘子，是指它满足：

(1) $\lambda_i\geq 0, \forall\, i=1,\cdots,k$;

(2) 真凸函数

$$f_0+\lambda_1 f_1+\cdots+\lambda_m f_m$$

的下确界有穷而且等于式 (5.3.4) 的最优值.

这样，一旦知道了凸规划问题 $(\mathcal{P})$ 的 Lagrange 乘子，就用不着先确定 $(\mathcal{P})$ 的可行解，然后在这些可行解集上求 f_0 的极小，而只要先求 $f_0+\lambda_1 f_1+\cdots+\lambda_m f_m$ 在整个$\mathbb{R}^n$ 上的极小，然后再从该极小集中去掉不满足约束条件的点.

定理 5.3.2 给定凸规划问题 $(\mathcal{P})$, 设 $(\lambda_1,\cdots,\lambda_m)^\tau$ 为 $(\mathcal{P})$ 的 Lagrange 乘子，并设

$$g=f_0+\lambda_1 f_1+\cdots+\lambda_m f_m.$$

如果 D 是 g 的极小集，并且

$$I=\{i\mid 1\leq i\leq k, \lambda_i=0\},\quad J=\{i\mid 1\leq i\leq m, i\notin I\}.$$

那么

$$D_0 = \{z \in D \mid f_j(z) = 0, \forall j \in J;\ f_i(z) \leq 0, \forall i \in I\}$$

是 $(\mathcal{P})$ 的最优解集.

证明 根据假设, $\inf g = \inf f$, 这里 $f = f_0 + \delta(\cdot|M)$, 并且 $\inf f$ 有穷. 设 z 为 $(\mathcal{P})$ 的任一可行解, 则

$$\lambda_i f_i(z) \leq 0, \quad i = 1, \cdots, m.$$

因此

$$g(z) = f_0(z) + \lambda_1 f_1(z) + \cdots + \lambda_m f_m(z) \leq f_0(z).$$

由此 $g(x) \leq f(x), \forall x \in \mathbb{R}^n$, 而且等号成立当且仅当 x 是 $(\mathcal{P})$ 的可行解, 且满足

$$\lambda_i f_i(x) = 0, \quad i = 1, \cdots, m.$$

这表明 f 的极小集包含在 g 的极小集中, 并且实际上就是 D_0. 但 f 的极小集就是 $(\mathcal{P})$ 的最优解集. ∎

推论 5.3.3 给定凸规划问题 $(\mathcal{P})$, 设 $(\lambda_1, \cdots, \lambda_m)^\tau$ 为 $(\mathcal{P})$ 的 Lagrange 乘子. 假定诸 f_i 都是闭的. 如果

$$g = f_0 + \lambda_1 f_1 + \cdots + \lambda_m f_m$$

的下确界 $\inf g$ 在唯一点 $\tilde{x}$ 处达到, 则 $\tilde{x}$ 是 $(\mathcal{P})$ 的唯一的最优解.

证明 从假设可知, g 和 $f = f_0 + \delta(\cdot|M)$ 均是闭的. 于是从定理 5.3.2 可知, 为了推论的结论成立, 只要证明 $(\mathcal{P})$ 至少有一个最优解. 但 g 的极小集是单点集, 故 g 没有回收方向, 即 $\operatorname{epi} g$ 不含任何 "水平" 方向的半直线 (见习题 2.4.4). 从定理 5.3.2 的证明知 $g \leq f$, 因此 $\operatorname{epi} f$ 也不能含任何 "水平" 方向的半直线, 从而 f 也没有回收方向. 最后从定理 5.2.2 知 f 的极小集非空. ∎

引理 5.3.4 设 M 为凸规划问题 $(\mathcal{P})$ 的可行解集，那么 $\forall x \in \mathbb{R}^n$,

$$\delta(x|M) = \sup\left\{\sum_{i=1}^{m}\lambda_i f_i(x) \mid \lambda \in \mathbb{R}_+^k \times \mathbb{R}^{m-k}\right\}, \quad (5.3.5)$$

其中 $\lambda = (\lambda_1, \cdots, \lambda_m)^\tau$.

证明 显然我们有

$$\begin{aligned}
&\sup\left\{\sum_{i=1}^{m}\lambda_i f_i(x) \mid \lambda \in \mathbb{R}_+^k \times \mathbb{R}^{m-k}\right\} \\
&= \sum_{i=1}^{k}\sup\{\lambda_i f_i(x) \mid \lambda_i \geq 0\} \\
&\quad + \sum_{j=k+1}^{m}\sup\{\lambda_j f_j(x) \mid \lambda_j \in \mathbb{R}\}.
\end{aligned}$$

但是对于 $1 \leq i \leq k$,

$$\sup\{\lambda_i f_i(x) \mid \lambda_i \geq 0\} = \begin{cases} 0, & \text{当 } f_i(x) \leq 0, \\ +\infty, & \text{当 } f_i(x) > 0; \end{cases}$$

而对于 $k+1 \leq j \leq m$,

$$\sup\{\lambda_j f_j(x) \mid \lambda_j \in \mathbb{R}\} = \begin{cases} 0, & \text{当 } f_j(x) = 0, \\ +\infty, & \text{当 } f_j(x) \neq 0. \end{cases}$$

因此由 M 和 $\delta(\cdot|M)$ 的定义知式 (5.3.5) 成立. ∎

一般地，我们定义函数 $L : \mathbb{R}^n \times \mathbb{R}_+^k \times \mathbb{R}^{m-k} \to (-\infty, +\infty]$ 如下：

$$L(x,\lambda) = f_0(x) + \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x).$$

L 称为凸规划问题 $(\mathcal{P})$ 的 Lagrange 函数. 令

$$\alpha = \sup_{\lambda \in \mathbb{R}^k_+ \times \mathbb{R}^{m-k}} \inf_{x \in \mathbb{R}^n} L(x,\lambda); \tag{5.3.6}$$

$$\beta = \inf_{x \in \mathbb{R}^n} \sup_{\lambda \in \mathbb{R}^k_+ \times \mathbb{R}^{m-k}} L(x,\lambda) \tag{5.3.7}$$

引理 5.3.5 由式 (5.3.7) 定义的 β 是凸规划问题 $(\mathcal{P})$ 的最优值，即

$$\inf f = \beta,$$

并且可行解 $\tilde{x} \in M$ 是 $(\mathcal{P})$ 的最优解 当且仅当

$$\sup_{\lambda \in \mathbb{R}^k_+ \times \mathbb{R}^{m-k}} L(\tilde{x},\lambda) = \beta.$$

证明 这是引理 5.3.4 的直接结果. ∎

从 α 和 β 的定义式 (5.3.6) 和 (5.3.7) 立即推出

$$\alpha \le \beta. \tag{5.3.8}$$

进而如果存在 $\tilde{\lambda} \in \mathbb{R}^k_+ \times \mathbb{R}^{m-k}$ 满足

$$\inf_{x \in \mathbb{R}^n} L(x,\tilde{\lambda}) = \beta, \tag{5.3.9}$$

则 $\alpha = \beta$. 这样，凸规划问题 $(\mathcal{P})$ 的 Lagrange 乘子就是指满足式 (5.3.9) 的向量 $\tilde{\lambda} \in \mathbb{R}^k_+ \times \mathbb{R}^{m-k}$, 并且 β 是有穷值.

因此我们得到如下结论：凸规划问题 $(\mathcal{P})$ 等价于

$$(\mathcal{P}) \qquad \sup_{\lambda \in \mathbb{R}^k_+ \times \mathbb{R}^{m-k}} L(x,\lambda) \to \min;$$

而问题 $(\mathcal{P})$ 的 Lagrangec 乘子 $\tilde{\lambda}$ 则必定是下列规划问题的解:

$$(\mathcal{P}^*) \qquad \begin{cases} \inf\limits_{x\in\mathbb{R}^n} L(x,\lambda) \to \max, \\ \lambda \in \mathbb{R}_+^k \times \mathbb{R}^{m-k}, \end{cases}$$

$(\mathcal{P}^*)$ 称为原问题 $(\mathcal{P})$ 的对偶规划问题. 注意在对偶规划问题 $(\mathcal{P}^*)$ 中 $\inf\limits_{x\in\mathbb{R}^n} L(x,\lambda)$ 是 λ 的凹函数而非凸函数. 但 $(\mathcal{P}^*)$ 也可以写成如下等价的凸规划问题形式:

$$(\mathcal{P}^*) \qquad \begin{cases} -\inf\limits_{x\in\mathbb{R}^n} L(x,\lambda) \to \min, \\ \lambda \in \mathbb{R}_+^k \times \mathbb{R}^{m-k}. \end{cases}$$

定理 5.3.6 给定凸规划问题 $(\mathcal{P})$, 那么下列两个命题等价:

(1) $\tilde{x}$ 是凸规划问题 $(\mathcal{P})$ 的最优解, 而 $\tilde{\lambda}$ 是 $(\mathcal{P})$ 的 lagrange 乘子;

(2) $(\tilde{x},\tilde{\lambda})$ 是 $(\mathcal{P})$ 的 Lagrange 函数 L 的鞍点, 即

$$\begin{cases} L(\tilde{x},\lambda) \le L(\tilde{x},\tilde{\lambda}) \le L(x,\tilde{\lambda}), \\ \forall\, x \in \mathbb{R}^n,\ \forall\,\lambda \in \mathbb{R}_+^k \times \mathbb{R}^{m-k}. \end{cases} \tag{5.3.10}$$

此外, 这时还有

$$\lambda_i f_i(\tilde{x}) = 0, \quad 1 \le i \le k. \tag{5.3.11}$$

证明 先设 (1) 成立. 于是由引理 5.3.2 和式 (5.3.9),

$$L(\tilde{x},\lambda) \le \beta \le L(x,\tilde{\lambda}), \quad \forall\, x \in \mathbb{R}^n, \forall\,\lambda \in \mathbb{R}_+^k \times \mathbb{R}^{m-k}.$$

但 $\beta = L(\tilde{x},\tilde{\lambda})$, 因此 (2) 成立.

再设 (2) 成立, 由式 (5.3.10) 得

$$\beta \leq \sup_{\lambda \in \mathbb{R}_+^k \times \mathbb{R}^{m-k}} L(\tilde{x}, \lambda) \leq L(\tilde{x}, \tilde{\lambda}) \leq \inf_{x \in \mathbb{R}^n} L(x, \tilde{\lambda}) \leq \alpha.$$

这样，由 $\alpha \leq \beta$ 得

$$\sup_{\lambda \in \mathbb{R}_+^k \times \mathbb{R}^{m-k}} L(\tilde{x}, \lambda) = L(\tilde{x}, \tilde{\lambda}) = \inf_{x \in \mathbb{R}^n} L(x, \tilde{\lambda}) = \beta.$$

因此由引理 5.3.5, $\tilde{x}$ 是 $(\mathcal{P})$ 的最优解；而且 $\tilde{\lambda}$ 满足 (5.3.9), 故 $\tilde{\lambda}$ 确实是 $(\mathcal{P})$ 的 Lagrange 乘子.

最后，由于 $\tilde{x} \in M$, 所以

$$L(\tilde{x}, \tilde{\lambda}) = f_0(\tilde{x}) + \sum_{i=1}^{k} \tilde{\lambda}_i f_i(\tilde{x}) = \beta = f_0(\tilde{x}).$$

由此可见

$$\sum_{i=1}^{k} \tilde{\lambda}_i f_i(\tilde{x}) = 0.$$

但 $\lambda_i \geq 0$, $f_i(\tilde{x}) \leq 0$, $i = 1, \cdots, k$, 因此 $\tilde{\lambda}_i f_i(\tilde{x}) = 0$, $i = 1, \cdots, k$. ∎

引进 Lagrange 乘子的好处是把原来的约束极值问题简化为无约束极值问题. 因此找出凸规划问题 $(\mathcal{P})$ 的 Lagrange 乘子对于凸规划问题 $(\mathcal{P})$ 具有重要意义. 但是在一般情况下，凸规划问题 $(\mathcal{P})$ 的 Lagrange 乘子不一定存在. 本节的后部分将研究凸规划问题 $(\mathcal{P})$ 的 Lagrange 乘子的存在性问题.

定理 5.3.7 设 $f_1, \cdots, f_k : \mathbb{R}^n \to (-\infty, +\infty]$ 为真凸函数， $g_1, \cdots, g_m$ 为 $\mathbb{R}^n$ 上的仿射函数. 记

$$N = \{x \in \mathbb{R}^n \mid g_j(x) = 0, \ 1 \leq j \leq m\}.$$

如果

$$F(x) \triangleq \max_{1\le i\le k} f_i(x) \ge 0, \quad \forall\, x \in N, \tag{5.3.12}$$

那么存在不同时为零的两个向量 $\lambda = (\lambda_1, \cdots, \lambda_k)^\tau \in I\!R^k_+$ 和 $\mu = (\mu_1, \cdots, \mu_m)^\tau \in I\!R^m$, 使得

$$\sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{j=1}^{m} \mu_j g_j(x) \ge 0, \quad \forall\, x \in \operatorname{dom} F, \tag{5.3.13}$$

这里显然 $\operatorname{dom} F = \bigcap\limits_{i=1}^{k} (\operatorname{dom} f_i)$. 此外，若记

$$g(x) = (g_1(x), \cdots, g_m(x))^\tau, \quad x \in I\!R^n,$$

则当

$$0 \in \operatorname{int}\,(g(\operatorname{dom} F))$$

时，式 (5.3.13) 也是式 (5.3.12) 的充分必要条件，并且这时 λ 为非零向量.

证明 定义映射 $T : \operatorname{dom} F \to I\!R^k \times I\!R^m$ 如下：

$$T(x) = (f_1(x), \cdots, f_k(x), g_1(x), \cdots, g_m(x))^\tau, \ x \in \operatorname{dom} F.$$

容易看出 $T(\operatorname{dom} F)$ 为 $I\!R^k \times I\!R^m$ 中的凸集；而且如果令

$$M = \{(y, z) \in I\!R^k \times I\!R^m \mid y < 0,\ z = 0\},$$

则从假设 (5.3.12) 可知 $M \cap T(\operatorname{dom} F) = \emptyset$. 因此由凸集分离定理，存在非零向量 $(\lambda, \mu) \in I\!R^k \times I\!R^m$, 使得

$$\begin{cases} \displaystyle\sum_{i=1}^{k} \lambda_i \eta_i \le \sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{j=1}^{m} \mu_j g_j(x), \\ \forall\, x \in \operatorname{dom} F, \forall\, y = (\eta_1, \cdots, \eta_m)^\tau < 0. \end{cases} \tag{5.3.14}$$

在式 (5.3.14) 中，让 $\eta_1, \cdots, \eta_k \to 0$, 即得式 (5.3.13) 成立. 由于式 (5.3.14) 中每一个 η_i 都是非负的，故必有 $\lambda_i \geq 0, 1 \leq i \leq k$.

现在假定 $0 \in \operatorname{int}(g(\operatorname{dom} F))$, 并且式 (5.3.13) 成立. 今证这时 $\lambda_1, \cdots, \lambda_k$ 不全为零. 事实上，否则我们有

$$\langle \mu, g(x) \rangle \geq 0, \quad \forall x \in \operatorname{dom} F,$$

即

$$\langle \mu, z \rangle \geq 0, \quad \forall z \in g(\operatorname{dom} F),$$

其中 $\mu = (\mu_1, \cdots, \mu_m)^\tau$. 但 $0 \in \operatorname{int}(g(\operatorname{dom} F))$, 因此上式仅当 $\mu = 0$ 时才能成立，这与 $(\lambda, \mu) \neq 0$ 相矛盾 (这里 $\lambda = (\lambda_1, \cdots, \lambda_k)^\tau$). 于是我们证明了 $\lambda \in I\!R^k_+, \lambda \neq 0$. 最后从式 (5.3.13) 可得

$$\lambda_1 f_1(x) + \cdots + \lambda_k f_k(x) \geq 0, \quad \forall x \in N.$$

由此可见对于每个 $x \in N$, 至少有一个标号 $i, 1 \leq i \leq k$, 使得 $f_i(x) \geq 0$, 从而式 (5.3.12) 成立. ∎

定理 5.3.8 给定凸规划问题 $(\mathcal{P})$. 假定

$$\tilde{x} \in \operatorname{int}\left[\bigcap_{i=0}^{k} \operatorname{dom} f_i\right],$$

其中 $g = (f_{k+1} \cdots, f_m)^\tau : I\!R^n \to I\!R^{m-k}$. 那么 $\tilde{x}$ 是问题 $(\mathcal{P})$ 的最优解当且仅当存在不同时为零的 $\lambda_0 \geq 0$ 和 $(\lambda_1, \cdots, \lambda_k)^\tau \in I\!R^k_+$, 以及 $(\lambda_{k+1}, \cdots, \lambda_m)^\tau \in I\!R^{m-k}$, 使得

$$0 \in \sum_{i=0}^{m} \lambda_i \partial f_i(\tilde{x}), \tag{5.3.15}$$

$$\lambda_i f_i(\tilde{x}) = 0, \quad 1 \leq i \leq k. \tag{5.3.16}$$

证明 不妨假定 $g: \mathbb{R}^n \to \mathbb{R}^{m-k}$ 是满射的,即 $g(\mathbb{R}^n) = \mathbb{R}^{m-k}$. 于是由 $g(\tilde{x}) = 0$ 可知

$$0 \in \operatorname{int}\Big[g\Big(\bigcap_{i=0}^{k} \operatorname{dom} f_i\Big)\Big].$$

不难看出 $\tilde{x}$ 是 $(\mathcal{P})$ 的最优解等价于 $\forall\, x : g(x) = 0$,

$$F(x) \triangleq \max_{1 \le i \le k} \Big\{f_0(x) - f_0(\tilde{x}), f_i(x)\Big\} \ge 0. \qquad (5.3.17)$$

于是依据定理 5.3.7, 式 (5.3.17) 成立的充分必要条件是存在 $(\lambda_0, \lambda_1, \cdots, \lambda_m)^\tau \in \mathbb{R}^{m+1}$, 并且 $\lambda_0 \ge 0, \cdots, \lambda_k \ge 0$ 不全为零, 使得

$$\lambda_0(f_0(x) - f_0(\tilde{x})) + \sum_{i=1}^{m} \lambda_i f_i(x) \ge 0, \quad \forall\, x \in \operatorname{dom} F, \qquad (5.3.18)$$

或者等价地

$$\lambda_0(f_0(x) - f_0(\tilde{x})) + \sum_{i=1}^{m} \lambda_i f_i(x) \ge 0, \quad \forall\, x \in \mathbb{R}^n.$$

上式表明 $\lambda_0(f_0(x) - f_0(\tilde{x})) + \sum\limits_{i=1}^{m} \lambda_i f_i(x)$ 在 $\tilde{x}$ 处达到其最小值. 但我们知道为此的充分必要条件是

$$0 \in \partial\Big(\lambda_0\Big(f_0 - f_0(\tilde{x})\Big) + \sum_{i=1}^{m} \lambda_i f_i\Big)(\tilde{x}) = \sum_{i=0}^{m} \lambda_i \partial f_i(\tilde{x}),$$

这里由于 $0 \in \operatorname{int}\Big[g(\bigcap\limits_{i=0}^{k} \operatorname{dom} f_i)\Big]$, 我们利用了关于次微分的 Rockafellar 定理. 至于式 (5.3.16) 则是式 (5.3.18) 的直接结论. ∎

注 如果 $f_0, f_1, \cdots, f_k$ 都在 $\tilde{x}$ 处连续并且有限，那么定理中的条件 $\tilde{x} \in \operatorname{int}\left[\bigcap_{i=0}^{k} \operatorname{dom} f_i\right]$ 自然满足.

定理 5.3.9 (Kuhn-Tucker) 在定理 5.3.8 的条件下，如果存在 $z \in I\!R^n$, 使得

$$f_i(z) < 0, 1 \leq i \leq k; \; f_j(z) = 0, k+1 \leq i \leq m, \tag{5.3.19}$$

那么式 (5.3.15) 中的 $\lambda_0 = 1$, 换句话说，在式 (5.3.19) 之下，$(\mathcal{P})$ 有最优解 $\tilde{x}$ 的充分必要条件是存在 $\lambda_1 \geq 0, \cdots, \lambda_k \geq 0, \lambda_{k+1}, \cdots, \lambda_m \in I\!R$, 使得

$$0 \in \partial f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i \partial f_i(\tilde{x}), \tag{5.3.20}$$

$$\lambda_i f_i(\tilde{x}) = 0, \quad 1 \leq i \leq k. \tag{5.3.21}$$

证明 只要证明必要性. 为此又只要证明 $\lambda_0 > 0$. 如果 $\lambda_0 = 0$, 则由式 (5.3.15),

$$0 \in \sum_{i=1}^{m} \lambda_i \partial f_i(\tilde{x}).$$

这表明凸函数 $\sum_{i=1}^{m} \lambda_i f_i(\tilde{x})$ 在 $\tilde{x}$ 取最小值零. 但 $\lambda_1, \cdots, \lambda_k$ 不全为零，从而与条件 (5.3.19) 相矛盾. ▮

定理 5.3.9 中的系数 $\lambda_1, \cdots, \lambda_m$ 就是问题 $(\mathcal{P})$ 的 Lagrange 乘子. 实际上为了问题 $(\mathcal{P})$ 的 Lagrange 乘子存在，定理 5.3.9 中所加的条件可以减弱. 我们有

定理 5.3.10 给定凸规划问题 $(\mathcal{P})$, 并设 $f_1, \cdots, f_k$ 中前 r 个为非仿射函数，而后 $k-r$ 个为仿射函数. 假定凸规划问题 $(\mathcal{P})$ 的最优值 $\alpha > -\infty$, 并且至少有一个 $z \in \operatorname{ri}(\operatorname{dom} f_0)$, 使得

$$\begin{cases} f_1(z) < 0, \cdots, f_r(z) < 0; \; f_{r+1}(z) \leq 0, \cdots, \\ f_k(z) \leq 0; \; f_{k+1}(z) = 0, \cdots, f_m(z) = 0. \end{cases} \tag{5.3.22}$$

那么凸规划问题 $(\mathcal{P})$ 的 Lagrange 乘子存在.

证明 我们先讨论无等式约束的情形，即 $k = m$. 于是不等方程组

$$\begin{cases} f_1(z) < 0, \cdots, f_r(z) < 0, \\ f_{r+1}(z) \le 0, \cdots, f_m(z) \le 0, \end{cases} \tag{5.3.23}$$

在 $\mathrm{ri}\,(\mathrm{dom}\, f_0)$ 中至少有一个解. 显然，不等方程组

$$\begin{cases} f_0(x) - \alpha < 0, f_1(x) < 0, \cdots, f_r(x) < 0, \\ f_{r+1}(z) \le 0, \cdots, f_m(z) \le 0, \end{cases}$$

在 $\mathrm{dom}\, f_0$ 中无解. 于是由定理 5.1.5, 存在不全为零的非负实数 $\lambda_0, \cdots, \lambda_m$, 使得

$$\lambda_0(f_0(x) - \alpha) + \sum_{i=1}^{m} \lambda_i f_i(x) \ge 0 \quad \forall\, x \in \mathrm{dom}\, f_0.$$

今证 $\lambda_0 > 0$. 事实上，若 $\lambda_0 = 0$, 则 $\lambda_1 f_1 + \cdots + \lambda_m f_m$ 在 $\mathrm{ri}\,(\mathrm{dom}\, f_0)$ 上非负，同时 $\lambda_1, \cdots, \lambda_r$ 中至少有一个是正的，这与式 (5.3.23) 在 $\mathrm{ri}\,(\mathrm{dom}\, f_0)$ 中有解相矛盾，故可以取 $\lambda_0 = 1$.

于是函数

$$g = f_0 + \lambda_1 f_1 + \cdots + \lambda_m f_m$$

满足 $g(x) \ge \alpha, \forall\, x \in \mathrm{dom}\, f_0$, 并且 $g(x) = +\infty, \forall\, x \notin \mathrm{dom}\, f_0$, 从而 $\inf g \ge \alpha$. 另一方面，对于任意可行解 $x \in M$, $g(x) \le f_0(x)$. 因此 $\inf g$ 不可能大于 f_0 在可行解集 M 上的下确界 α. 这样 $\inf g = \alpha$, 并且 $\lambda_1, \cdots, \lambda_m$ 是问题 $(\mathcal{P})$ 的 Lagrange 乘子.

现在我们讨论出现等式约束的情形，即 $k < m$, 这时依问题 $(\mathcal{P})$ 的定义， $f_{k+1}, \cdots, f_m$ 是仿射函数. 显然每个约束 $f_i(x) = 0$ 可以用两个不等式约束

$$f_i(x) \le 0, \quad (-f_i)(x) \le 0,$$

替代而得到一个新的等价的凸规划问题 $(\mathcal{P}')$, 这时 $(\mathcal{P}')$ 仅有不等式约束. 于是上面已经证明了的部分适用于问题 $(\mathcal{P}')$, 得出结论：存在非负实数 $\lambda_1, \cdots, \lambda_k, \lambda'_{k+1}, \cdots, \lambda'_m, \lambda''_{k+1}, \cdots, \lambda''_m$, 使得函数

$$f_0 + \sum_{i=1}^{k} \lambda_i f_i + \sum_{j=k+1}^{m} \lambda'_j f_j + \sum_{j=k+1}^{m} \lambda''_j f_j$$

的下确界有穷，并且等于 $(\mathcal{P}')$ 的最优值. 但 $(\mathcal{P}')$ 与 $(\mathcal{P})$ 的最优值相等，因此，如果令 $\lambda_j = \lambda'_j - \lambda''_j, j = k+1, \cdots, \lambda_m$, 则我们得到 $(\mathcal{P})$ 的 Lagrange 乘子 $(\lambda_1, \cdots, \lambda_m)$. ∎

推论 5.3.11 给定凸规划问题 $(\mathcal{P})$, $k = m$, 即问题 $(\mathcal{P})$ 无等式约束. 假定 $(\mathcal{P})$ 的最优值不等于 $-\infty$, 并且不等方程组

$$f_1(x) < 0, \cdots, f_m(x) < 0,$$

在 $\operatorname{dom} f_0$ 中有解. 那么问题 $(\mathcal{P})$ 有 Lagrange 乘子.

证明 设 $x \in \operatorname{dom} f_0$ 满足 $f_i(x) < 0, 1 \le i \le m$. 并且任取一点 $y \in \operatorname{ri}(\operatorname{dom} f_0)$, 从而依本节开头的假设 (5.3.1), $y \in \operatorname{ri}(\operatorname{dom} f_i), 1 \le i \le m$. 于是当 $\lambda > 0$ 充分小时, $z(\lambda) = (1-\lambda)x + \lambda y$ 满足 $f_i(z(\lambda)) < 0, 1 \le i \le m$, 并且 $z(\lambda) \in \operatorname{ri}(\operatorname{dom} f_0)$. 这样，定理 5.3.10 的条件满足. ∎

推论 5.3.12 设凸规划问题 $(\mathcal{P})$ 仅有线性约束，即

$$f_i(x) = \langle a_i, x \rangle - \alpha_i, \quad i = 1, \cdots, m.$$

假定 $(\mathcal{P})$ 的最优值不等于 $-\infty$, 并且 $(\mathcal{P})$ 在 $\mathrm{ri}\,(\mathrm{dom}\, f_0)$ 中至少有一个可行解，那么 $(\mathcal{P})$ 有 Lagrange 乘子.

一般说来，凸规划问题 $(\mathcal{P})$ 不一定存在 Lagrange 乘子. 例如，设 $\mathrm{dom}\, f_0 = I\!R^2, f_0(\xi_1,\xi_2) = \xi_1, f_1(\xi_1,\xi_2) = \xi_1^2 - \xi_2, f_2(\xi_1,\xi_2) = \xi_2, k = 2 = m$. 在目前情况下，不等方程组

$$f_1(\xi_1,\xi_2) \le 0, \quad f_2(\xi_1,\xi_2) \le 0,$$

仅有一解 $x = (0,0)^\tau$. 因此 $(0,0)^\tau$ 是该凸规划问题的唯一最优解，并且其最优值为 0. 然而，如果 (λ_1,λ_2) 为其 Lagrange 乘子，则 $\lambda_1 \ge 0, \lambda_2 \ge 0$, 并且

$$\begin{aligned} 0 &\le f_0(\xi_1,\xi_2) + \lambda_1 f_1(\xi_1,\xi_2) + \lambda_2 f_2(\xi_1,\xi_2) \\ &= \xi_1 + (\lambda_1 - \lambda_2)\xi_2 + \lambda_2 \xi_1^2, \quad \forall (\xi_1,\xi_2)^\tau \in I\!R^2. \end{aligned}$$

这显然是不可能的. 在这个例子里，定理 5.3.10 的假设不满足，因为没有一个元 $(\xi_1,\xi_2)^\tau \in I\!R^2$ 能满足 $f_1(\xi_1,\xi_2) < 0$, $f_2(\xi_1,\xi_2) \le 0$.

下一个例子说明，当上面定理中关于相对内部的条件不满足时也可能没有 Lagrange 乘子. 取

$$f_0(\xi_1,\xi_2) = \xi_1, \mathrm{dom}\, f_0 = \{(\xi_1,\xi_2)^\tau \mid \xi_1^2 - \xi_2 \le 0\};$$

$$f_1(\xi_1,\xi_2) = \xi_2;\ k = 0, m = 1.$$

在这个凸规划问题中， $f_1(\xi_1,\xi_2) = 0$ 在 $\mathrm{ri}\,(\mathrm{dom}\, f_0)$ 中无解. 而 $x = (0,0)^\tau$ 仍然是其唯一的最优解，0 是其最优值. 这时 Lagrange 乘子是单个实数 $\lambda_1 \ge 0$ 并且必须满足

$$\begin{cases} 0 \le f_0(\xi_1,\xi_2) + \lambda_1 f_1(\xi_1,\xi_2) = \xi_1 + \lambda_1 \xi_2, \\ \qquad \forall (\xi_1,\xi_2)^\tau \in \mathrm{dom}\, f_0. \end{cases}$$

但容易看出，这样的 $\lambda_1 \geq 0$ 是不存在的.

在结束本节之前，我们讨论一下凸规划问题 $(\mathcal{P})$ 的摄动问题，即讨论约束条件的变化的影响. 给定凸规划问题 $(\mathcal{P})$ 以及向量 $a = (a_1, \cdots, a_m)^\tau \in \mathbb{R}^m$. 我们考虑如下新的凸规划问题 $(\mathcal{P}_a)$:

$$(\mathcal{P}_a) \qquad \begin{cases} f_0(x) \to \min, \\ f_i(x) + a_i \leq 0, \quad 1 \leq i \leq k, \\ f_j(x) + a_j = 0, \quad k+1 \leq j \leq m, \end{cases}$$

这里 $f_0, \cdots, f_m$ 是凸规划问题 $(\mathcal{P})$ 中的函数. 设 $v(a)$ 为问题 $(\mathcal{P}_a)$ 的最优值. 于是 $v(\cdot)$ 是 $\mathbb{R}^m$ 上的函数.

定理 5.3.13 设 $v(a)$ 为凸规划问题 $(\mathcal{P}_a)$ 的最优值，那么 $v(\cdot)$ 是 $\mathbb{R}^m$ 上的凸函数.

证明 我们只要证明 $v(\cdot)$ 在 $\operatorname{dom} v = \{a \in \mathbb{R}^m \mid v(a) < +\infty\}$ 中满足凸性不等式. 设 $a, b \in \operatorname{dom} v$, $a = (a_1, \cdots, a_m)^\tau$, $b = (b_1, \cdots, b_m)^\tau$, 并且 $v(a), v(b) \neq -\infty$. 于是 $\forall \varepsilon > 0$, $\exists x, y \in \mathbb{R}^n$, 使得

$$\begin{cases} f_0(x) \leq v(a) + \varepsilon, \\ f_i(x) \leq -a_i, \quad 1 \leq i \leq k, \\ f_j(x) = -a_j, \quad k+1 \leq j \leq m; \\ f_0(y) \leq v(b) + \varepsilon, \\ f_i(y) \leq -b_i, \quad 1 \leq i \leq k, \\ f_j(y) = -b_j, \quad k+1 \leq j \leq m. \end{cases}$$

由于 $f_1, \cdots, f_k$ 为凸函数，而 $f_{k+1}, \cdots, f_m$ 为仿射函数，故对于任意 $0 < \lambda < 1$, 有

$$f_0((1-\lambda)x + \lambda y) \leq (1-\lambda) f_0(x) + \lambda f_0(y)$$

$$
\begin{aligned}
&\le (1-\lambda)v(a)+\lambda v(b)+\varepsilon,\\
f_i((1-\lambda)x+\lambda y) &\le (1-\lambda)f_i(x)+\lambda f_i(y)\\
&\le -(1-\lambda)a_i-\lambda b_i, \quad 1\le i\le k,\\
f_j((1-\lambda)x+\lambda y) &= (1-\lambda)f_j(x)+\lambda f_j(y)\\
&= -(1-\lambda)a_i-\lambda b_i, \quad k+1\le j\le m.
\end{aligned}
$$

因此 $(1-\lambda)x+\lambda y$ 满足问题 $(\mathcal{P}_{(1-\lambda)a+\lambda b})$ 的约束条件，从而

$$
\begin{aligned}
v((1-\lambda)a+\lambda b) &\le f_0((1-\lambda)x+\lambda y)\\
&\le (1-\lambda)v(a)+\lambda v(b)+\varepsilon.
\end{aligned}
$$

但 $\varepsilon>0$ 是任意的，所以

$$
v((1-\lambda)a+\lambda b)\le (1-\lambda)v(a)+\lambda v(b).
$$

当 $v(a)$ 或 $v(b)=-\infty$ 时，用 $-q$ (q 为任意正数) 代替上面的 $v(a)+\varepsilon$ 或 $v(b)+\varepsilon$，上面的凸性不等式仍然成立. ∎

定理 5.3.14 设凸规划问题 $(\mathcal{P}_{\tilde{a}})$ 的 Lagrange 乘子存在，并且设 S 为问题 $(\mathcal{P}_{\tilde{a}})$ 的 Lagrange 乘子全体. 那么 $v(\cdot)$ 是 $\mathbb{R}^n$ 上的真凸函数，它在 $\tilde{a}$ 处次可微，并且

$$
\partial v(\tilde{a})=S.
$$

证明 设 $\lambda\in N$，即 λ 是问题 $(\mathcal{P}_{\tilde{a}})$ 的 Lagrange 乘子. 于是

$$
\begin{aligned}
v(\tilde{a}) &= \inf_{x\in\mathbb{R}^n}\{f_0(x)+\sum_{i=1}^m\lambda_i(f_i(x)+\tilde{a}_i)\}\\
&= \inf_{x\in\mathbb{R}^n}\{f_0(x)+\sum_{i=1}^m\lambda_i f_i(x)\}+\sum_{i=1}^m\lambda_i\tilde{a}_i.
\end{aligned}
$$

由于 $\lambda_1, \cdots, \lambda_m \geq 0$, 如果记

$$S_x = \{a \in \mathbb{R}^m \mid f_i(x) + a_i \leq 0, 1 \leq i \leq k;$$
$$f_j(x) + a_j = 0, k+1 \leq j \leq m\}, x \in \mathbb{R}^n,$$

$$M_a = \{x \in \mathbb{R}^n \mid f_i(x) + a_i \leq 0, 1 \leq i \leq k;$$
$$f_j(x) + a_j = 0, k+1 \leq j \leq m\}, a \in \mathbb{R}^m,$$

则

$$\begin{aligned} v(\tilde{a}) &= \inf_{x \in \mathbb{R}^n} \inf_{a \in S_x} \{f_0(x) - \langle \lambda, a \rangle\} + \langle \lambda, \tilde{a} \rangle \\ &= \inf_{a \in \mathbb{R}^m} \inf_{x \in M_a} \{f_0(x) - \langle \lambda, a \rangle\} + \langle \lambda, \tilde{a} \rangle \\ &= \inf_{a \in \mathbb{R}^m} \{v(a) - \langle \lambda, a \rangle\} + \langle \lambda, \tilde{a} \rangle. \end{aligned}$$

由此即得

$$v(a) \geq v(\tilde{a}) + \langle \lambda, a - \tilde{a} \rangle, \quad \forall a \in \mathbb{R}^m,$$

即 $\lambda \in \partial v(\tilde{a})$.

把上述推理倒推过去即知 $\partial v(\tilde{a})$ 中的每一个元均是问题 $(\mathcal{P}_{\tilde{a}})$ 的 Lagrange 乘子. ∎

习 题 5.3

5.3.1 下列形式的规划问题称为二次规划问题:

$$\begin{cases} \frac{1}{2}\langle Ax, x \rangle + \langle c, x \rangle \to \min, \\ Bx \geq 0, \ x \geq 0. \end{cases}$$

这里 $x \in \mathbb{R}^n, c \in \mathbb{R}^n$, A 为 $n \times n$ 对称正定矩阵, B 为 $m \times n$ 矩阵. 试讨论它的对偶问题及其 Lagrange 乘子.

5.3.2 设 $f: \mathbb{R}^n \to (-\infty, +\infty]$ 为真凸函数, $K \subset \mathbb{R}^n$ 为非空闭凸锥. 试指出极值问题

$$f(x) \to \min, \quad x \in K$$

的对偶问题为

$$-f^*(x^*) \to \max, \quad x^* \in K^o,$$

其中 K^o 表示 K 的极化锥. 提示：先证明

$$\delta(x|K) = \sup\{\langle x, x^*\rangle \mid x^* \in K^o\},$$

然后定义 Lagrange 函数 $L : \mathbb{R}^n \times K^o \to (-\infty, +\infty]$, $L(x, x^*) = f(x) + \langle x, x^*\rangle$.

5.4 线性规划问题

当凸规划问题 $(\mathcal{P})$ 中出现的所有函数 f_i 都是线性函数时，称为线性规划问题. 于是一般线性规划问题可以写成：

$$(\mathcal{L}) \qquad \begin{cases} \sum_{i=1}^{n} c_i\xi_i \to \min, \\ \sum_{j=1}^{n} a_{ij}\xi_j \geq b_i, \quad i = 1, \cdots, m, \\ \xi_j \geq 0, \quad j = 1, \cdots, n. \end{cases}$$

令 $c = (c_1, \cdots, c_n)^\tau$, $b = (b_1, \cdots, b_m)^\tau$, $x = (\xi_1, \cdots, \xi_n)^\tau$, $A = (a_{ij})$, 则上述线性规划问题可以简写为

$$(\mathcal{L}) \qquad \begin{cases} \langle c, x\rangle \to \min, \\ Ax \geq b, \\ x \geq 0. \end{cases}$$

大量的实际问题可以归结为线性规划问题. 它可作如下的经济解释：设 $x = (\xi_1, \cdots, \xi_n)^\tau$ 表示某生产过程中所需要的 n 种原料的投入量， $c = (c_1, \cdots, c_n)^\tau$ 表示这 n 种原料的

单位价格. 于是

$$\langle c, x\rangle = \sum_{i=1}^{n} c_i \xi_i$$

就是所有投入量的总价值，即生产成本. 又设 $b = (b_1, \cdots, b_m)^T$ 表示这一生产过程最终的 m 种产品的产出量，矩阵 A 的元 a_{ij} 表示投入第 j 种原料对产出第 i 种产品的消耗系数. 于是 $Ax \geq b$ 意指要求投入量为 x 时产出量应不少于 b. 相应的线性规划问题 $(\mathcal{L})$ 就是在指定的产出量限制下要求投入的成本最小.

实际规划问题中某些不等号可能是反向的，也可能有等式约束，但我们总可以把它们化成上述的标准形式.

线性规划问题 $(\mathcal{L})$ 完全可以没有最优解，这里可能出现两种情况：其一是可能某些不等式约束不相容，以致相应的可行解集 $K \triangleq \{x \in \mathbb{R}^n \mid x \geq 0, Ax \geq b\} = \emptyset$. 这时我们称问题 $(\mathcal{L})$ 是不可行的. 其二是可能目标函数 $f(x) = \langle c, x\rangle$ 在 K 上可以取任意大的值，以致达不到最大值. 这时我们称问题 $(\mathcal{L})$ 是无界的.

线性规划问题 $(\mathcal{L})$ 是一般凸规划问题的特殊情形，因此可以应用上一节的一般凸规划问题的理论. 对应于问题 $(\mathcal{L})$ 的 Lagrange 函数为

$$\begin{aligned} L(x, \mu, y) &= \langle c, x\rangle - \langle \mu, x\rangle + \langle y, b - Ax\rangle \\ &= \sum_{i=1}^{n}(c_i - \mu_i - \sum_{j=1}^{m} a_{ij}\eta_j)\xi_i + \sum_{j=1}^{m} \eta_j b_j, \end{aligned} \tag{5.4.1}$$

这里为了使下面的讨论对称起见，我们用 μ, y 表示 $(\mathcal{L})$ 的 Lagrange 乘子， $\mu \in \mathbb{R}^n, y \in \mathbb{R}^m$, 并且为记号简单起见，不同维数空间中的欧氏内积采用了同一符号 $\langle\cdot,\cdot\rangle$. 从上节知道， $(\mathcal{L})$ 的 Lagrange 乘子应该是对偶问题

$$\inf_{x\in\mathbb{R}^n} L(x,\mu,y) \to \max$$

的解. 但由 (5.4.1) 容易推出

$$\inf_{x\in\mathbb{R}^n} L(x,\mu,y) = \begin{cases} \sum_{j=1}^{m} b_j\eta_j, & \text{当 } \mu = c - A^\tau y, \\ -\infty, & \text{其余处}. \end{cases}$$

因此 y 是下列对偶线性规划问题 $(\mathcal{L}^*)$ 的解:

$$(\mathcal{L}^*) \qquad \begin{cases} \langle b, y\rangle \to \max, \\ A^\tau y \le c, \\ y \ge 0. \end{cases}$$

而 μ 由 $\mu = c - A^\tau y$ 给出. $(\mathcal{L}^*)$ 也是一个线性规划问题, 它也有对偶问题 $(\mathcal{L}^{**})$. 但不难看出

$$(\mathcal{L}^{**}) = (\mathcal{L}),$$

即 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 是互为对偶的线性规划问题.

设 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 的可行解集分别为

$$K = \{x \in \mathbb{R}^n \mid x \ge 0, Ax \ge b\},$$

$$K_* = \{y \in \mathbb{R}^m \mid y \ge 0, A^\tau y \le c\}.$$

根据一般凸规划的理论, 容易得到下列线性规划问题的基本定理.

定理 5.4.1 设 $(\mathcal{L})$ 为线性规划问题, 那么下列几个命题彼此等价:

(1) $(\mathcal{L})$ 有最优解 $\tilde{x}$;

(2) $(\mathcal{L})$ 的最优值有穷;

(3) $(\mathcal{L}^*)$ 有最优解 $\tilde{y}$;

(4) $(\mathcal{L}^*)$ 的最优值有穷;

(5) $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 的可行解集 K 和 K_* 都非空. 在上述任何一种情况下都有

$$\inf\{\langle c, x\rangle \mid x \in K\} = \sup\{\langle b, y\rangle \mid y \in K_*\},$$

并且对于相应的最优解 $\tilde{x}$ 和 $\tilde{y}$, 有

$$\tilde{\eta}_i(\sum_{j=1}^{n} a_{ij}\tilde{\xi}_j - b_i) = 0, \quad i = 1, \cdots, m,$$

$$\tilde{\xi}_j(\sum_{i=1}^{m} a_{ij}\tilde{\eta}_i - c_j) = 0, \quad j = 1, \cdots, n.$$

其中$\tilde{x} = (\tilde{\xi}_1, \cdots, \tilde{\xi}_n)^{\tau}$, $\tilde{y} = (\tilde{\eta}_1, \cdots, \tilde{\eta}_m)^{\tau}$.

证明 (1)$\Longleftrightarrow$(2) 和 (3)$\Longleftrightarrow$ (4) 是显然的. 为证 (2)$\Longleftrightarrow$ (3), 注意 $(\mathcal{L})$ 的最优值有穷表明 $K \neq \emptyset$, 从而依定理 5.3.10, $(\mathcal{L})$ 存在 Lagrange 乘子, 即 $(\mathcal{L}^*)$ 有最优解. (4)$\Longleftrightarrow$ (1) 与 (2)$\Longleftrightarrow$ (3) 是相仿的. 因此 (1)–(4) 彼此等价. 现在设 (1) 成立, 从而 (3) 成立, 于是立即得出 (5) 成立. 最后设 (5) 成立, 则

$$\inf\{\langle c, x\rangle \mid x \in K\} = \inf_{x \in \mathbb{R}^n} \sup_{\mu \in \mathbb{R}^n_+, y \in \mathbb{R}^m_+} L(x, \mu, y) = \beta < \infty,$$

$$\sup\{\langle b, y\rangle \mid y \in K_*\} = \sup_{\mu \in \mathbb{R}^n_+, y \in \mathbb{R}^m_+} \inf_{x \in \mathbb{R}^n} L(x, \mu, y) = \alpha > -\infty.$$

但 $\alpha \leq \beta$, 因此

$$-\infty < \sup\{\langle b, y\rangle \mid y \in K_*\} \leq \inf\{\langle c, x\rangle \mid x \in K\} < \infty,$$

即 (2) 和 (4) 成立. 其余结论可由定理 5.3.6 得到. ∎

线性规划问题现在是运筹学的一个重要的独立分支. 上面我们从凸规划的一般理论对线性规划问题作了讨论. 为了读者阅读方便起见，这里我们对线性规划问题的一般理论和方法作出相对独立的叙述.

我们把线性规划问题 $(\mathcal{L})$ 及其对偶问题 $(\mathcal{L}^*)$ 重新列出于下:

$$\text{原问题}(\mathcal{L}) \quad \begin{cases} f(x) \triangleq \langle c, x \rangle \to \min, \\ Ax \geq b, \\ x \geq 0; \end{cases}$$

$$\text{对偶问题}(\mathcal{L}^*) \quad \begin{cases} g(y) \triangleq \langle b, y \rangle \to \max, \\ A^\tau y \leq c, \\ y \geq 0. \end{cases}$$

这里 $c \in \mathbb{R}^n, b \in \mathbb{R}^m, x \in \mathbb{R}^n, y \in \mathbb{R}^m$, 而 A 是 $m \times n$ 实矩阵. K 和 K_* 分别表示 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 的可行解集:

$$K = \{x \in \mathbb{R}^n \mid x \geq 0, Ax \geq b\},$$

$$K_* = \{y \in \mathbb{R}^m \mid y \geq 0, A^\tau y \leq c\}.$$

引理 5.4.2 线性规划问题 $(\mathcal{L})$ 的可行解集 K 是 $\mathbb{R}^n$ 中的一个不含任何直线的多面形集合.

证明 问题 $(\mathcal{L})$ 中的每一个约束都是一个线性不等式，从而确定 $\mathbb{R}^n$ 中一闭半空间. 可行解集 K 是有穷多个这样的闭半空间之交，从而是一个多面形集合. 此外，不等式约束 $x \geq 0$ 意味着 K 位于非负象限，因此不含任何直线. ▮

定理 5.4.3 假定线性规划问题 $(\mathcal{L})$ 的目标函数 $f(x) = \langle c, x \rangle$ 在 K 上下有界，并且 $K \neq \emptyset$, 那么线性规划问题 $(\mathcal{L})$

至少有一个最优解. 此外，最优解中至少有一个是 K 的极点.

证明 直接从引理 5.4.2 和定理 1.7.6 得到. ∎

定理 5.4.3 不仅告诉我们最优解什么时候存在，而且还指出了寻找最优解的一种可能的方法：即计算目标函数 $f(x) = \langle c, x \rangle$ 在 K 的每个极点上的值，然后在其中选出达到最小值的点. 这个办法在一些简单情形确实是可行的 (见例 5.4.1)，但遗憾的是在许多实际应用中基本上是不可行的. 这是因为工业过程中一个典型的线性规划问题可能涉及到上千个变量和约束条件，从而即使采用高速计算机，要找出 K 的所有极点也得花费大量的时间.

寻找最优解而同时又能避免这种大量运算的一种方法是所谓的"单纯形法"，其思想大致如下：

(1) 找出 K 的一个极点；

(2) 考察 K 的所有在 x 处相会的棱边. 如果目标函数 $f(x)$ 沿着这些棱边都不减少，则 x 是最优解.

(3) 另一方面，若目标函数 $f(x)$ 沿着这些棱边中的某几条会减少，则选择减少得最多的一条棱边，并把极点移至该棱边的一端.

(4) 在新的极点处重复上述 (2) 开始的步骤. 由于 f 在每一步中都减少，故我们不可能两次通过同一个极点. 但 K 的极点总数是有限的，因此这一过程经过有穷步 (当然希望步数越少越好) 后总会给出最优解 (如果最优解存在的话). 如果线性规划问题是无界的，则必然会在 (3) 这一步中目标函数 $f(x)$ 沿着某条棱边无限减少.

例 5.4.1 考虑如下线性规划问题：

$$(\mathcal{L})\quad \begin{cases} f(\xi_1,\xi_2)=2\xi_1-3\xi_2\to\min, \\ \xi_1\le 30,\ \xi_2\le 20,\ \xi_1+2\xi_2\le 52, \\ \xi_1\ge 0,\ \xi_2\ge 0. \end{cases}$$

从图 5.4.1 看出可行解集 K 是一个五边形. 通过求解适当的不等式对可以找出所有的极点.

极点 (ξ_1,ξ_2)	(0,0)	(30,0)	(30,11)	(12,20)	(0,20)
$f(\xi_1,\xi_2)$	0	-60	-27	36	60

计算目标函数在每个极点处的值，得到最小值 -60, 相应的最优解为 (30,0).

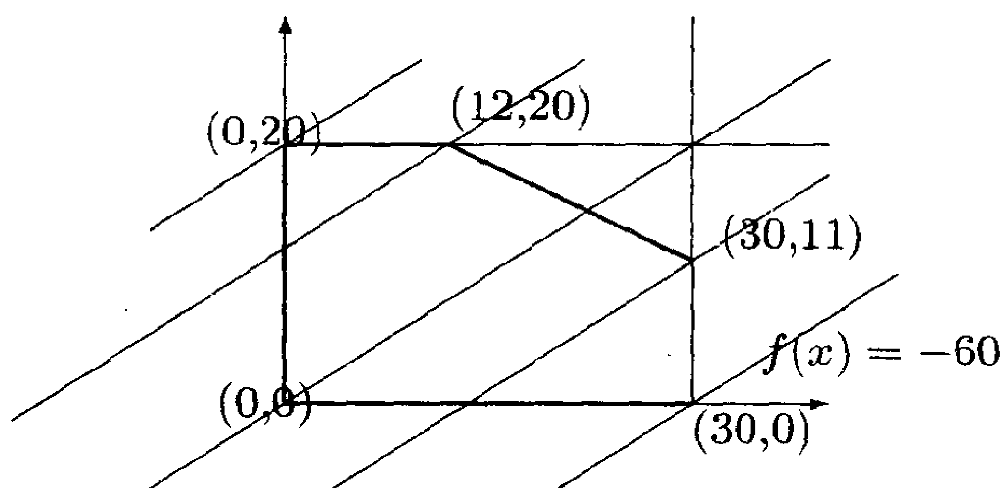


图 5.4.1

当线性规划问题仅含两个变量时，则可以用几何方法或图解法 (见图 5.4.1). 很明显，通过点 (30,0) 的直线 $f(x)=-60$ 对应于在可行解集上的最小值.

现在我们讨论线性规划问题的原问题 $(\mathcal{L})$ 和对偶问题 $(\mathcal{L}^*)$ 之间的关系.

定理 5.4.4 设 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 分别为线性规划问题的原问题及其对偶问题，相应的可行解集分别为 K 和 K_*, 那么

(1) $x \in K, y \in K_* \Longrightarrow f(x) \geq g(y)$;

(2) 如果对于 $\tilde{x} \in K, \tilde{y} \in K_*$, 有 $f(\tilde{x}) = g(\tilde{y})$, 则 $\tilde{x}$ 是 $(\mathcal{L})$ 的最优解，而 $\tilde{y}$ 是 $(\mathcal{L}_*)$ 的最优解.

证明 (1) 若 $x \in K, y \in K_*$, 从定义可知

$$\begin{aligned} f(x) &= \langle c, x\rangle \geq \langle A^T y, x\rangle \\ &= \langle y, Ax\rangle \geq \langle y, b\rangle = g(y). \end{aligned}$$

(2) 若对于 $\tilde{x} \in K, \tilde{y} \in K_*$, 有 $f(\tilde{x}) = g(\tilde{y})$, 则对于任意 $x \in K$, 根据 (1) 有

$$f(x) \geq g(\tilde{y}) = f(\tilde{x}).$$

从而 $\tilde{x}$ 是 $(\mathcal{L})$ 的最优解. 类似地，对于任意 $y \in K_*$, 有

$$g(y) \leq f(\tilde{x}) = g(\tilde{y}).$$

因此 $\tilde{y}$ 是 $(\mathcal{L}_*)$ 的最优解. ∎

在证明线性规划问题的基本的对偶定理之前，我们先证明两个引理.

引理 5.4.5 (Farkas引理) 下列两个命题二择一成立:

(1) 方程

$$Ax = b \tag{5.4.2}$$

有非负解;

(2) 不等式组

$$A^T y \geq 0, \quad \langle b, y\rangle < 0, \tag{5.4.3}$$

有解.

证明 先假定 (1) 成立，即式 (5.4.2) 有一非负解 x. 如果式 (5.4.3) 有解 y. 那么

$$0 > \langle b, y \rangle = \langle Ax, y \rangle = \langle x, A^{\tau} y \rangle.$$

但 $x \geq 0, A^{\tau} y \geq 0$, 从而上式不可能成立. 因此上述两个命题不可能同时成立.

现在设式 (5.4.2) 没有非负解. 于是单点集合 $\{b\}$ 和集合

$$S = \{z \in \mathbb{R}^m \mid z = Ax, x \geq 0\}$$

不相交. 显然 S 是一个闭凸锥. 于是依凸集分离定理 1.5.6, 存在非零向量 $z_0 \in \mathbb{R}^m$ 和实数 α 使得超平面 $H(z_0, \alpha)$ 把 $\{b\}$ 与 S 强分离, 而且由于 $0 \in S$, 我们可以假定

$$\begin{cases} \langle b, z_0 \rangle < \alpha < 0, \\ \langle z, z_0 \rangle > \alpha, \quad \forall z \in S. \end{cases} \tag{5.4.4}$$

不难看出, 与 $H(z_0, \alpha)$ 相平行的超平面 $H(z_0, 0)$ 是 S 在原点处的承托超平面, 即

$$\langle z, z_0 \rangle \geq 0, \quad \forall z \in S.$$

(这是因为, 如果有某个 $z \in S$ 使得 $\langle z, z_0 \rangle < 0$, 则对于充分大的正数 k, $\langle kz, z_0 \rangle > \alpha$ 就不可能成立, 然而 $kz \in S, \forall k > 0$).

这样我们证明了

$$\langle Ax, z_0 \rangle \geq 0, \quad \forall x \geq 0,$$

即

$$\langle x, A^{\tau} z_0 \rangle \geq 0, \quad \forall x \geq 0. \tag{5.4.5}$$

我们断定: 式 (5.4.5) 蕴涵着 $A^{\tau} z_0 \geq 0$. 事实上, 如果 $A^{\tau} z_0$ 的第 j 个分量为负数, 则 $\langle e_j, A^{\tau} z_0 \rangle < 0$, (这里 e_j 表示 $\mathbb{R}^n$ 中第 j 个分量为 1, 而其余分量为 0 的规范基向量), 这是不

可能的. 因此我们有

$$\langle b, z_0 \rangle < 0, \quad A^{\tau} z_0 \geq 0.$$

从而向量 z_0 是 (5.4.3) 的解. ∎

引理 5.4.6 下列两个命题二择一成立:

(1) 方程

$$Ax \geq b, \tag{5.4.6}$$

有非负解;

(2) 不等式组

$$A^{\tau} y \leq 0, \quad \langle b, y \rangle > 0, \tag{5.4.7}$$

有非负解.

证明 仿引理 5.4.5 的证明容易推出式 (5.4.6) 和式 (5.4.7) 不可能同时有非负解. 现在我们设式 (5.4.6) 没有非负解. 这意味着对于 $z \in \mathbb{R}^m$, 方程

$$-Ax + z = -b,$$

没有非负解 $x \geq$, $z \geq 0$. 上式可以写成如下等价的形式:

$$[-A, I_m] \begin{bmatrix} x \\ z \end{bmatrix} = -b, \tag{5.4.8}$$

其中 I_m 为 m 阶单位矩阵. 由于式 (5.4.8) 没有非负解 $[x, z]^{\tau}$, 从引理 5.4.5 知存在向量 $y \in \mathbb{R}^m$ 使得

$$\begin{bmatrix} -A^{\tau} \\ I_m \end{bmatrix} y \geq 0, \quad \langle b, y \rangle > 0.$$

但 $\begin{bmatrix} -A^{\tau} \\ I_m \end{bmatrix} y \geq 0$ 意味着 $A^{\tau} y \leq 0$ 和 $y \geq 0$, 因此我们找到了 (5.4.7) 的非负解 y. ∎

定理 5.4.7 对于线性规划问题 $(\mathcal{L})$ 及其对偶问题 $(\mathcal{L}^*)$, 我们有

(1) 如果 K 和 K_* 均非空，则 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 都有最优解；进而设 $\tilde{x}$ 和 $\tilde{y}$ 分别为 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 的最优解，则 $f(\tilde{x}) = g(\tilde{y})$.

(2) 如果 K 和 K_* 中有一个是空集，则 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 都没有最优解.

(3) 如果 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 之一有最优解 $\tilde{x}$ 或 $\tilde{y}$, 则另一个问题也有最优解，并且$f(\tilde{x}) = g(\tilde{y})$.

证明 设 $x \in K, y \in K_*$, 即 x 和 y 分别是方程

$$Ax \geq b \tag{5.4.9}$$

和

$$A^{\tau} y \leq c \tag{5.4.10}$$

的非负解. 如果我们能找到式 (5.4.9) 和式 (5.4.10) 的解 x 和 y 使之满足

$$\langle c, x\rangle - \langle y, b\rangle \leq 0, \tag{5.4.11}$$

则定理的 (1) 得证. 事实上，从定理 5.4.4 之 (1) 可知 $\langle c, x\rangle \geq \langle b, y\rangle$, 故若式 (5.4.11) 成立，必有 $\langle c, x\rangle = \langle b, y\rangle$. 于是再从定理 5.4.4 之 (2) 可知 x 和 y 是最优解.

现在我们把式 (5.4.9)–(5.4.11) 联合写成等价的形式：

$$\begin{bmatrix} A & 0 \\ 0 & -A^{\tau} \\ -c^{\tau} & b^{\tau} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \geq \begin{bmatrix} b \\ -c \\ 0 \end{bmatrix}, \tag{5.4.12}$$

这里 0 不加区分地表示适当阶数的零矩阵. 假定式 (5.4.9)—(5.4.11) 没有非负解，即式 (5.4.12) 没有非负解. 于是从引理 5.4.6 知道存在向量 $z \in \mathbb{R}^m, z \geq 0$ 和 $w \in \mathbb{R}^n, w \geq 0$,

以及一实数 $\alpha \geq 0$ 使得

$$\begin{cases} \begin{bmatrix} A^\tau & 0 & -c \\ 0 & -A & b \end{bmatrix} \begin{bmatrix} z \\ w \\ \alpha \end{bmatrix} \leq 0, \\ \left\langle \begin{bmatrix} z \\ w \\ \alpha \end{bmatrix}, \begin{bmatrix} b \\ -c \\ 0 \end{bmatrix} \right\rangle > 0. \end{cases} \tag{5.4.13}$$

式 (5.4.13) 等价于

$$\begin{cases} A^\tau z \leq \alpha c, \\ Aw \geq \alpha b, \\ \langle b, z \rangle - \langle w, c \rangle > 0. \end{cases} \tag{5.4.14}$$

α 不可能等于 0, 因为否则的话, 从式 (5.4.14) 的前两个不等式以及式 (5.4.9) 和式 (5.4.10), 可知对于 $x \in K$ 和 $y \in K_*$,

$$\begin{cases} 0 \geq \langle A^\tau z, x \rangle = \langle z, Ax \rangle \geq \langle z, b \rangle, \\ 0 \leq \langle Aw, y \rangle = \langle w, A^\tau y \rangle \leq \langle w, c \rangle, \end{cases}$$

这与式 (5.4.14) 的第三个不等式相抵触. 于是 $\alpha > 0$.

因此从式 (5.4.14) 的前两个不等式可知 $w/\alpha \in K, z/\alpha \in K_*$, 从而根据定理 5.4.4 之 (1),

$$\langle w/\alpha, c \rangle \geq \langle z/\alpha, b \rangle,$$

即

$$\langle w, c \rangle \geq \langle z, b \rangle.$$

这与式 (5.4.14) 的第三个不等式相矛盾.

这样我们得出结论: 问题 (5.4.9)—(5.4.11) 有非负解 (x, y), $x \in K$, $y \in K_*$, 并且从上面所述, 它们正好是问题 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 的最优解.

为证 (2), 比如我们假定 $K=\emptyset$, 于是不等式 $Ax\geq b$ 没有非负解. 根据引理 5.4.6, 必有向量 $z\geq 0$, 使得

$$A^{\tau}z\leq 0,\quad \langle b,z\rangle>0. \tag{5.4.15}$$

如果 $K_*\neq\emptyset$, 即存在向量 $y\in\mathbb{R}^m, y\geq 0$ 使得

$$A^{\tau}y\leq c,$$

则从式 (5.4.15) 可见 $y+\lambda z\in K_*, \forall\lambda>0$. 由于 $\langle b,z\rangle>0$, 可知内积

$$\langle b,y+\lambda z\rangle=\langle b,y\rangle+\lambda\langle b,z\rangle$$

可以取任意大的正数值. 这就是说目标函数 $g(y)=\langle b,y\rangle$ 在 K_* 上上无界，即对偶问题 $(\mathcal{L}^*)$ 没有最优解.

最后，(3) 容易从 (1) 和 (2) 推出. 事实上，比如说 $(\mathcal{L})$ 有最优解 $\tilde{x}$, 则根据 (2), $K_*\neq\emptyset$. 再依据 (1), $(\mathcal{L}^*)$ 有最优解 $\tilde{y}$, 并且 $f(\tilde{x})=g(\tilde{y})$. ∎

下一个定理在线性规划理论中通常称为平衡定理.

定理 5.4.8 设 K 和 K_* 分别为线性规划问题 $(\mathcal{L})$ 及其对偶问题 $(\mathcal{L}^*)$ 的可行解集. 那么为了 $x\in K$ 和 $y\in K_*$ 分别为 $(\mathcal{L})$ 和 $(\mathcal{L}^*)$ 的最优解，必须且只须

$$\xi_j=0,\quad \text{如果 } (A^{\tau}y)_j<c_j, \tag{5.4.16}$$

$$\eta_i=0,\quad \text{如果 } (Ax)_i>b_i, \tag{5.4.17}$$

其中 $x=(\xi_1,\cdots,\xi_n)^{\tau}, y=(\eta_1,\cdots,\eta_m)^{\tau}, c=(c_1,\cdots,c_n)^{\tau}$, $b=(b_1,\cdots,b_n)^{\tau}$.

证明 首先假定式 (5.4.16) 和式 (5.4.17) 成立. 于是从 $y\in K_*$ 以及式 (5.4.16) 得到

$$f(x) = \langle c, x\rangle = \sum_{j=1}^{n} c_j\xi_j$$

$$= \sum_{j=1}^{n}(A^\tau y)_j\xi_j = \langle A^\tau y, x\rangle.$$

类似地，从 $x \in K$ 和式 (5.4.17) 得到

$$g(y) = \langle b, y\rangle = \sum_{i=1}^{m} b_i\eta_i$$

$$= \sum_{i=1}^{m}(Ax)_i\eta_i = \langle y, Ax\rangle.$$

但 $\langle y, Ax\rangle = \langle A^\tau y, x\rangle$，因此 $f(x) = g(y)$. 从而根据定理 5.4.4 之 (2), x 和 y 是最优解.

反之，如果 x 和 y 是最优解，于是从定理 5.4.7 有 $\langle c, x\rangle = \langle b, y\rangle$. 另一方面，

$$\langle c, x\rangle = \sum_{j=1}^{n} c_j\xi_j \geq \sum_{j=1}^{n}(A^\tau y)_j\xi_j$$

$$= \sum_{i=1}^{m}(Ax)_i\eta_i \geq \sum_{i=1}^{m} b_i\eta_i = \langle b, y\rangle.$$

因此

$$\langle c - A^\tau y, x\rangle = 0, \quad \langle y, Ax - b\rangle = 0. \tag{5.4.18}$$

于是考虑到 $x \geq 0, c - A^\tau \geq 0, y \geq 0, Ax - b \geq 0$, 从式 (5.4.18) 得到

$$[c_j - (A^\tau y)_j]\xi_j = 0, \quad j = 1, \cdots, n;$$

$$\eta_i[(Ax)_i - b_i] = 0, \quad i = 1, \cdots, m.$$

据此立即得到条件 (5.4.16) 和 (5.4.17). ∎

习 题 5.4

5.4.1 求解下列线性规划问题:

(1) $\begin{cases} 5\xi_1 + 4\xi_2 \to \min, \\ 2\xi_1 + 5\xi_2 \geq 10,\ \ 3\xi_1 + \xi_2 \geq 6, \\ \xi_1 + 7\xi_2 \geq 7,\ \ \xi_1 \geq 0, \xi_2 \geq 0. \end{cases}$

(2) $\begin{cases} 2\eta_1 + \eta_2 \to \max, \\ \eta_1 - \eta_2 \leq 3,\ \ -\eta_1 + 2\eta_2 \leq -4,\ \ \eta_1 \geq 0, \eta_2 \geq 0. \end{cases}$

(3) $\begin{cases} 3\xi_1 + 2\xi_2 \to \min, \\ 2\xi_1 - \xi_2 \geq 4,\ \ -\xi_1 + \xi_2 \geq 4,\ \ \xi_1 \geq 0, \xi_2 \geq 0. \end{cases}$

(4) $\begin{cases} 8\eta_1 + 7\eta_2 \to \max, \\ 2\eta_1 + \eta_2 \leq 32,\ \ \eta_1 + \eta_2 \leq 18, \\ \eta_1 + 3\eta_2 \leq 24,\ \ \eta_1 \geq 0, \eta_2 \geq 0. \end{cases}$

5.4.2 考虑如下线性规划问题 $(\mathcal{L})$:

$$\begin{cases} 80\xi_1 + 120\xi_2 \to \min, \\ 5\xi_1 + 5\xi_2 \geq 9,\ \ 2\xi_1 + 4\xi_2 \geq 5,\ \ \xi_1 \geq 0, \xi_2 \geq 0. \end{cases}$$

(1) 求解问题 $(\mathcal{L})$,

(2) 叙述对偶问题 $(\mathcal{L}^*)$,

(3) 求解对偶问题 $(\mathcal{L}^*)$.

5.4.3 如果线性规划问题 $(\mathcal{L})$ 是无界的, 试证对偶问题 $(\mathcal{L}^*)$ 是不可行的.

5.4.4 举例说明线性规划问题 $(\mathcal{L})$ 和其对偶问题 $(\mathcal{L}^*)$ 可能两者都是不可行的.

5.4.5 试给出对偶问题 $(\mathcal{L})$ 的经济解释.

5.4.6 设 $z_1, \cdots, z_m \in \mathbb{R}^n$. 试证不等方程组

$$\langle z_i, x\rangle < 0, \quad i = 1, \cdots, m$$

有解的充分必要条件是: 不存在不全为零的实数 $\lambda_1, \cdots, \lambda_m \geq 0$ 使得 $\lambda_1 z_1 + \cdots + \lambda_m z_m = 0$.

5.4.7 设 $z_0, z_1, \cdots, z_m \in \mathbb{R}^n$. 试证不等方程组

$$\langle z_0, x\rangle < 0,\ \langle z_i, x\rangle \le 0, \quad i = 1, \cdots, m$$

有解的充分必要条件是：不存在实数 $\lambda_1, \cdots, \lambda_m \ge 0$ 使得 $z_0 + \lambda_1 z_1 + \cdots + \lambda_m z_m = 0$.

5.4.8 设 $z_1, \cdots, z_m \in \mathbb{R}^n, \alpha_1, \cdots, \alpha_m \in \mathbb{R}$. 试证不等方程组

$$\langle z_i, x\rangle < \alpha_i, \quad i = 1, \cdots, m$$

有解的充分必要条件是：不存在不全为零的实数 $\lambda_1, \cdots, \lambda_m \ge 0$ 使得

$$\lambda_1 z_1 + \cdots + \lambda_m z_m = 0, \quad \lambda_1 \alpha_1 + \cdots + \lambda_m \alpha_m \le 0.$$

5.5 一般规划问题*

在 5.3 节中讨论的凸规划问题有一定的局限性，例如，所涉及的函数都是凸函数，而且在等式约束中出现的函数还必须是仿射函数. 因此，即使当目标函数和等式约束中的函数连续可微但非凸，或者等式约束中含非仿射的凸函数时就遇到了困难. 本章最后这一节中我们介绍有关非凸数学规划的一些结果. 不过要注意的是，尽管这里涉及到一些非凸函数，但我们用的基本的数学工具仍然是凸分析理论.

设给定 $\mathbb{R}^n$ 上的函数 $f, g_1, \cdots, g_m, h_1, \cdots, h_k$, 考虑如下数学规划问题：

$$\begin{cases} f(x) \to \min. \\ g_i(x) \le 0, \quad i = 1, \cdots, m, \\ h_j(x) = 0, \quad j = 1, \cdots, k. \end{cases} \tag{5.5.1}$$

定义不等式约束集和等式约束集如下：

$$M_i = \{x \in \mathbb{R}^n \mid g_i(x) \le 0\}, \quad i = 1, \cdots, m,$$

$$M_{m+1} = \{x \in I\!R^n \mid h_j(x) = 0, j = 1, \cdots, k\},$$

并令

$$M = \bigcap_{i=1}^{m+1} M_i.$$

于是问题 (5.5.1) 就是：求 $x_0 \in M$ 使得 $f(x)$ 在 M 上于 x_0 处达到极小，即

$$f(x_0) = \min_{x \in M} f(x).$$

对于一般的数学规划问题，我们要考虑局部极小. 我们称 x_0 是问题 (5.5.1) 的局部极小点，是指存在 x_0 的某个邻域 $U(x_0)$(例如以 x_0 为中心 ε 为半径的开球), 使得

$$f(x_0) = \min_{x \in M \cap U(x_0)} f(x).$$

设 $f : I\!R^n \to (-\infty, +\infty]$, 任取 $x \in I\!R^n$ 使得 $f(x) < +\infty$. 我们称非零向量 $y \in I\!R^n$ 所确定的方向是 f 在 x 的减小方向，是指存在两个正数 ε_0 和 α 以及 y 的一个邻域 $U(y)$, 使得

$$f(x + \varepsilon z) \le f(x) - \varepsilon\alpha, \quad \forall 0 < \varepsilon < \varepsilon_0, \forall z \in U(y). \tag{5.5.2}$$

注意，当 $z \in U(y)$ 时，$U(y)$ 也是 z 的一个邻域；并且对于任意 $\lambda > 0$, $\lambda U(y)$ 是 λy 的一个邻域. 由此可见，f 在 x 处的减小方向全体形成一个开锥，叫做 f 在 x 处的减小方向锥.

不等式约束集通常是含内点的集合. 设 M 是 $I\!R^n$ 中含内点的集合，非零向量 $y \in I\!R^n$ 所确定的方向叫做点 $x \in I\!R^n$ 关于 M 的容许方向，是指存在 $\varepsilon_0 > 0$ 和 y 的一个邻域 $U(y)$, 使得

$$x + \varepsilon z \in M, \quad \forall 0 < \varepsilon < \varepsilon_0, \forall z \in U(y). \tag{5.5.3}$$

同样地，x 处关于 M 的容许方向全体构成一个开锥，叫做 x 处关于 M 的容许方向锥.

一般说来，等式约束集是比 $\mathbb{R}^n$ 更低维数的集合. 于是设 M 是 $\mathbb{R}^n$ 中的低维集合，$y \in \mathbb{R}^n$ 叫做 M 在 x 处的切方向，是指存在 $\varepsilon_0 > 0$ 使得对于任意 $\varepsilon \in (0, \varepsilon_0)$, 都有一个 $x(\varepsilon) \in M$, 并且

$$\|x(\varepsilon) - x - \varepsilon y\| = o(\varepsilon).$$

M 在 x 处的切方向全体形成一个锥，叫做 M 在 x 处的切方向锥. 一般说来，这个切方向锥既非开也非闭.

给定函数 $f : \mathbb{R}^n \to (-\infty, +\infty]$ 和 $\mathbb{R}^n$ 中的约束集 $M_1, \cdots, M_{m+1}$, 其中 $M_1, \cdots, M_m$ 是 $\mathbb{R}^n$ 中含内点的集合. 假定 f 在 $M = \bigcap_{i=1}^{m+1} M_i$ 中某点 x_0 达到局部极小. 这就是说，f 从这点 x_0 出发沿任何符合约束条件的方向都不能再减小了. 因此，f 在 x_0 的减小方向中没有一个既在 x_0 关于 M_i 的容许方向锥 K_i 中 $(i = 1, \cdots, m)$, 又在 M_{m+1} 在 x_0 的切方向锥 K_{m+1} 中. 下一个定理就是这种直观分析结果的严格证明.

定理 5.5.1 设 f 和 $M_1, \cdots, M_{m+1}$ 如上. 设 K_i 为 x_0 关于 M_i 的容许方向锥 $(i = 1, \cdots, m)$, K_{m+1} 为 M_{m+1} 在 x_0 的切方向锥，而 K_0 为 f 在 x_0 的减小方向锥. 假定所有这些锥都是凸的. 如果 f 在 $M = \bigcap_{i=1}^{m+1} M_i$ 上于 x_0 处达到局部极小，则必有

$$\bigcap_{i=0}^{m+1} M_i = \emptyset. \tag{5.5.4}$$

证明 设不然，存在 $y \in \bigcap_{i=0}^{m+1} M_i$. 于是根据 K_i 的定义，存在 y 的一个邻域 U, 以及两个数 $\alpha > 0$ 和 $\varepsilon_0 > 0$, 使得

$$x_0+\varepsilon z\in\bigcap_{i=1}^m M_i,\quad \forall\, 0<\varepsilon<\varepsilon_0,\ \forall\, z\in U,$$

并且

$$f(x_0+\varepsilon z)\le f(x_0)-\varepsilon\alpha,\quad \forall\, 0<\varepsilon<\varepsilon_0,\ \forall\, z\in U.$$

另一方面，由于 $y\in K_{m+1}$，存在 $x(\varepsilon)=x_0+\varepsilon y+r(\varepsilon)\in M_{m+1}$，$\forall\varepsilon\in(0,\varepsilon_0)$，其中 $\|r(\varepsilon)\|=o(\varepsilon)$. 因此可以取 $\varepsilon_1, 0<\varepsilon_1\le\varepsilon_0$，使得

$$z(\varepsilon)\triangleq y+\frac{r(\varepsilon)}{\varepsilon}\in U,\quad \forall\, 0<\varepsilon<\varepsilon_1.$$

这样，当 $0<\varepsilon<\varepsilon_1$ 时，$x_0+\varepsilon z(\varepsilon)\in\bigcap_{i=1}^{m+1}M_i$，同时还满足

$$f(x_0+\varepsilon z(\varepsilon))\le f(x_0)-\varepsilon\alpha<f(x_0).$$

这与 f 在 M 上于 x_0 达到局部极小相矛盾. ∎

下面我们进一步分析条件 (5.5.4).

引理 5.5.2 设 $K_1,\cdots,K_m$ 为 $\mathbb{R}^n$ 中的凸锥. 假定

$$\mathrm{ri}\,K_1\cap\cdots\cap\mathrm{ri}\,K_m\neq\emptyset,\tag{5.5.5}$$

那么

$$\Big(\bigcap_{i=1}^m K_i\Big)^*=\sum_{i=1}^m K_i^*,\tag{5.5.6}$$

其中 K^* 表示 K 的对偶锥，即

$$K^*=\{x^*\in\mathbb{R}^n\mid\langle x,x^*\rangle\ge 0,\forall\, x\in K\}.$$

证明 我们只要证明

$$\Big(\bigcap_{i=1}^m K_i\Big)^o=\sum_{i=1}^m K_i^o,\tag{5.5.7}$$

这里 K^o 表示 K 的极化锥，即 K 的负对偶锥. 不妨设 $0 \in K_i, i=1,\cdots,m$. 于是

$$K_i^o = \{x^* \in I\!R^n \mid \langle x, x^*\rangle \leq 0, \forall\, x \in K_i\} = \partial\delta_{K_i}(0),$$

这里 $\delta_K(x) \triangleq \delta(x|K)$. 由于

$$\delta\Big(x\Big|\bigcap_{i=1}^{m} K_i\Big) = \sum_{i=1}^{m}\delta(x|K_i),$$

不难看出，式 (5.5.7) 等价于

$$\partial\Big(\sum_{i=1}^{m}\delta_{K_i}\Big)(0) = \sum_{i=1}^{m}\partial\delta_{K_i}(0).$$

但在条件 (5.5.5) 之下，这是 Rockafellar 定理的结果. ▮

定理 5.5.3 设 $K_0, K_1, \cdots, K_{m+1}$ 为 $I\!R^n$ 中 $m+1$ 个非空凸锥，并且 $K_0, \cdots, K_m$ 还是开的. 那么

$$\bigcap_{i=0}^{m+1} K_i = \emptyset \tag{5.5.8}$$

等价于存在不全为零的向量 $x_i^* \in K_i^*, i=0,\cdots,m+1$, 使得

$$x_0^* + \cdots + x_{m+1}^* = 0. \tag{5.5.9}$$

证明 首先假定 (5.5.8) 成立，并设 $K \triangleq \bigcap_{i=0}^{m} K_i \neq \emptyset$. 于是根据引理 5.5.2,

$$K^* = \Big(\bigcap_{i=0}^{m} K_i\Big)^* = \sum_{i=0}^{m} K_i^*.$$

从式 (5.5.8) 有 $K \cap K_{m+1} = \emptyset$. 因此依凸集分离定理，存在非零向量 $x^* \in I\!R^n$, 使得

$$\langle x^*, y\rangle \geq 0 \geq \langle x^*, x\rangle, \quad \forall\, x \in K_{m+1},\ \forall\, y \in K.$$

这表明 $x^* \in K^*$, 因此存在 $x_i^* \in K_i^*, i=0,\cdots,m$, 使得 $x^* = x_0^* + \cdots + x_m^*$. 今取 $x_{m+1}^* = -x^*$, 则显然 $x_{m+1}^* \in K_{m+1}^*$, 并且式 (5.5.9) 成立.

现在设 $\bigcap\limits_{i=0}^{m} K_i = \emptyset$. 于是存在 j, $1 \le j < m$, 使得

$$K = \bigcap_{i=0}^{j} K_i \neq \emptyset, \quad \bigcap_{i=0}^{j+1} K_i = \emptyset.$$

用 j 代替上面的 m, 可知存在 $x_i^* \in K_i^*, i=0,1,\cdots,j+1$, 使得 $x_{j+1}^* \neq 0$, 并且

$$x_0^* + \cdots + x_{j+1}^* = 0.$$

然后取 $x_{j+2}^* = \cdots = x_{m+1}^* = 0$, 可知式 (5.5.9) 成立.

最后假定式 (5.5.9) 成立, 但式 (5.5.8) 不成立. 于是存在 $x_0 \in K_i, i=0,1,\cdots,m+1$. 由于诸 x_i^* 不全为零, 于是比如说, 存在某个指标 $j, 0 \le j \le m$, 使得 $x_j^* \neq 0$. 此外, 由于 K_j 是开集, 故存在 $r>0$ 使得 $x_0 + B(0,r) \subset K_j$, 这里 $B(0,r)$ 为 $\mathbb{R}^n$ 中的开球. 考虑到 $x_j^* \in K_j^*$, 我们有

$$x_j^*(x_0+z) \ge 0, \quad \forall z \in B(0,r),$$

即

$$|x_j^*(z)| \le x_j^*(x_0), \quad \forall z \in B(0,r).$$

由此可见若 $x_j^*(x_0) = 0$, 则 $x_j^* = 0$, 这不可能. 从而 $x_j^*(x_0) > 0$. 于是

$$0 = x_0^*(x_0) + \cdots + x_{m+1}^*(x_0) \ge x_j^*(x_0) > 0,$$

得出矛盾. 因此式 (5.5.8) 成立. ∎

下一个定理是上述定理的直接结论.

定理 5.5.4 设 $f: \mathbb{R}^n \to (-\infty, -+\infty]$ 在 $M = \bigcap\limits_{i=1}^{m+1} M_i$ 上于 x_0 处达到局部极小, 这里 $M_1,\cdots,M_m$ 是含

内点的不等式约束集， M_{m+1} 是等式约束集. 假定 f 在 x_0 处的减小方向锥 K_0, x_0 关于 M_i 的容许方向锥 $K_i(i=1,\cdots,m)$, 以及 M_{m+1} 在 x_0 处的切方向锥 K_{m+1} 都是凸的. 那么必存在不全为零的 $m+2$ 个向量 $x_i^*\in K_i^*, i=0,1,\cdots,m+1$, 使得

$$x_0^*+x_1^*+\cdots+x_m^*+x_{m+1}^*=0. \tag{5.5.10}$$

为了把上述定理应用到具体的规划问题，我们需要研究具体情形下减小方向锥，容许方向锥和切方向锥的具体形式.

引理 5.5.5 设 y 是函数 f 在 x_0 处的减小方向. 如果 $f'_+(x_0;y)$ 存在，则$f'_+(x_0;y)<0$.

证明 由定义直接推出. ∎

定理 5.5.6 设 f 是 $\mathbb{R}^n$ 上的真凸函数, 并且在 x_0 的某邻域中连续. 那么 f 在 x_0 处的减小方向锥 K 是凸的, 并且

$$K=\{y\in\mathbb{R}^n \mid f'_+(x_0;y)<0\}.$$

证明 对于任意 $y\in\mathbb{R}^n$, $f'_+(x_0;y)$ 的存在性和凸性是 f 的凸性的结果. 为了完成证明，依据引理 5.5.5, 只需证明: $f'_+(x_0;y)<0\Longrightarrow y$ 是 f 在 x_0 处的减小方向. 依据 $f'_+(x_0;y)$ 的定义，存在 $\varepsilon_0>0$, 使得

$$\delta\triangleq f(x_0)-f(x_0+\varepsilon_0 y)>0.$$

根据 f 的连续性假设，存在 y 的一个邻域 U 使得

$$|f(x_0+\varepsilon_0 z)-f(x_0+\varepsilon_0 y)|\le\delta/2,\quad \forall\, z\in U.$$

从 f 的凸性可知，对于 $0<\varepsilon<\varepsilon_0$,

$$
\begin{aligned}
f(x_0+\varepsilon z) &\le (1-\varepsilon/\varepsilon_0)f(x_0)+(\varepsilon/\varepsilon_0)f(x_0+\varepsilon_0 z)\\
&\le f(x_0)-\frac{\delta}{2}\frac{\varepsilon}{\varepsilon_0}, \quad \forall\, z\in U.
\end{aligned}
$$

由此可知 y 是 f 在 x_0 处的减小方向. ∎

注意定理 5.5.6 中 f 在 x_0 的某邻域中连续的假设当 $\operatorname{dom} f$ 包含 x_0 的某个邻域时自然成立.

定理 5.5.7 设 $\mathbb{R}^n$ 上的函数 f 在 $x_0\in\mathbb{R}^n$ 处可微. 那么 f 在 x_0 处的减小方向锥 K 是凸的，并且

$$
K=\{y\in\mathbb{R}^n \mid \langle y, f'(x_0)\rangle<0\},
$$

其中 $f'(x_0)=\nabla f(x_0)$.

证明 由 f 在 x_0 处可微，对于任意 $y\in\mathbb{R}^n$, 我们有

$$
f(x_0+z)=f(x_0)+\langle f'(x_0),z\rangle+r(z),
$$

其中 $\|r(z)\|=o(\|z\|)$. 假定 $\langle f'(x_0),y\rangle\triangleq-2\alpha<0$. 于是存在 $\varepsilon_0>0$, 使得

$$
\frac{\|r(\varepsilon z)\|}{\varepsilon}<\frac{\alpha}{2}, \quad \forall\, 0<\varepsilon<\varepsilon_0,\ \forall\, z\in B(y,\rho),
$$

其中 $\rho\triangleq\alpha/2\|f'(x_0)\|$. 于是当 $z\in B(y,\rho)$ 时,

$$
\begin{aligned}
\langle f'(x_0),z\rangle &= \langle f'(x_0),y\rangle+\langle f'(x_0),z-y\rangle\\
&\le -2\alpha+\alpha/2=-3\alpha/2.
\end{aligned}
$$

这样我们证明了

$$
f(x_0+\varepsilon z)\le f(x_0)-\alpha\varepsilon, \quad \forall\, 0<\varepsilon<\varepsilon_0,\ \forall\, z\in B(y,\rho),
$$

即 y 是 f 在 x_0 处的减小方向. ∎

现在我们讨论容许方向锥.

定理 5.5.8 设 M 是 $\mathbb{R}^n$ 中的凸集. 设 K 是 x_0 关于 M 的容许方向锥. 那么

$$K = \cup\{\lambda(\operatorname{int} M - x_0) \mid \lambda > 0\}. \tag{5.5.11}$$

证明 依定义, $y \in K$ 当且仅当 $\exists \varepsilon_0 > 0$ 和 $r > 0$, 使得

$$x_0 + \varepsilon z \in M, \quad \forall\, 0 < \varepsilon < \varepsilon_0,\ \forall\, z \in B(y, r),$$

或等价地

$$x_0 + \varepsilon y + \varepsilon u \in M, \quad \forall\, 0 < \varepsilon < \varepsilon_0,\ \forall\, u \in B(0, r).$$

这就是说,

$$B(x_0 + \varepsilon y, \varepsilon r) \subset M.$$

从而 $x_0 + \varepsilon y \in \operatorname{int} M$, 即 $y \in \frac{1}{\varepsilon}(\operatorname{int} M - x_0)$. 然后取 $\lambda = 1/\varepsilon$, 并考虑到 K 的锥性, 可知式 (5.5.11) 成立. ∎

下面考虑 M 是不等式约束集的情形.

定理 5.5.9 (1) 设 K_0 是函数 $f: \mathbb{R}^n \to (-\infty, +\infty]$ 在 x_0 处的减小方向锥. 如果记

$$M = \{x \in \mathbb{R}^n \mid f(x) \le f(x_0)\},$$

并设 K_1 是 x_0 关于 M 的容许方向锥. 那么 $K_0 \subset K_1$.

(2) 如果 f 在 x_0 处沿任意方向 z 的方向导数 $f'_+(x_0; z)$ 存在, 并且 $f'_+(x_0; z)$ 是 $z \in \mathbb{R}^n$ 的凸函数. 又设存在一元 $\tilde{y} \in \mathbb{R}^n$, 使得

$$f'_+(x_0; \tilde{y}) < 0,$$

那么$K_0 = K_1 = \{z \in \mathbb{R}^n \mid f'_+(x_0; z) < 0\}$.

证明 (1) 设 $y \in K_0$, 于是依据定义, 存在 y 的一个邻域 $U(y)$, 以及两个正数 ε_0 和 α, 使得

$$f(x_0+\varepsilon z) \le f(x_0)-\varepsilon\alpha < f(x_0), \quad \forall\, 0<\varepsilon<\varepsilon_0,\ \forall\, z\in U(y).$$

而这正是说：$0<\varepsilon<\varepsilon_0$, $z\in U(y) \Longrightarrow x_0+\varepsilon z\in M$. 从而依据容许方向的定义，$y\in K_1$.

(2) 设 $y\in K_1$, 于是依据定义，存在 $\varepsilon_0>0$ 使得 $x_0+\varepsilon y\in M$, $\forall\, 0<\varepsilon<\varepsilon_0$, 即

$$f(x_0+\varepsilon y)\le f(x_0), \quad 0<\varepsilon<\varepsilon_0.$$

由此可见 $f'_+(x_0;y)\le 0$. 由于容许方向锥 K_1 是开的，故 y 有一邻域 $U(y)\subset K_1$. 于是可以取 $r>0$, 使得 $y(r)=y+r(y-\tilde{y})\in U(y)$. 这样 $y(r)\in K_1$, 从而 $f'_+(x_0;y(r))\le 0$. 最后利用 $f'_+(x_0;\cdot)$ 的凸性，我们有

$$\begin{aligned} f'_+(x_0;y) &= f'_+\Big(x_0;\frac{1}{1+r}y(r)+\frac{r}{1+r}\tilde{y}\Big)\\ &\le \frac{1}{1+r}f'_+(x_0;y(r))+\frac{r}{1+r}f'_+(x_0;\tilde{y})\\ &\le \frac{r}{1+r}f'_+(x_0;\tilde{y})<0. \end{aligned}$$

于是依据定理 5.5.6, $y\in K_0$. ∎

推论 5.5.10 设 K_0 和 K_1 如定理 5.5.9. 如果下列两条件之一成立，则 $K_0=K_1$:

(1) f 是 $\mathbb{R}^n$ 上凸连续函数，并且存在一元 $\tilde{x}\in\mathbb{R}^n$, 使得 $f(\tilde{x})<f(x_0)$;

(2) f 在 x_0 处可微，并且$f'(x_0)\ne 0$.

现在我们来讨论一般等式约束集的切方向锥. 给定映射 $F:\mathbb{R}^n\to\mathbb{R}^m$, 我们知道 F 在 $x_0\in\mathbb{R}^n$ 的导数 $F'(x_0)$ 是指 $\mathbb{R}^n\to\mathbb{R}^m$ 的一个线性变换，满足

$$F(x)=F(x_0)+F'(x_0)(x-x_0)+o(\|x-x_0\|). \quad (5.5.12)$$

定理 5.5.11 给定映射 $F: \mathbb{R}^n \to \mathbb{R}^m$. 设 F 在 x_0 的一个邻域 U 中可微, 并且 $F(x_0)=0$. 又设 $F'(x)$ 在 U 中连续并且 $F'(x_0)$ 是满秩的, 即 $F'(x_0)\mathbb{R}^n=\mathbb{R}^m$. 那么等式约束集

$$M=\{x\in\mathbb{R}^n \mid F(x)=0\}$$

在 x_0 处的切方向锥 K 是 $\mathbb{R}^n$ 的子空间

$$K=\{y\in\mathbb{R}^n \mid F'(x_0)y=0\}=N(F'(x_0)). \quad (5.5.13)$$

证明 首先设 $y\in K$. 于是依据定义, 存在 $\varepsilon_0>0$, 使得对于每个 $\varepsilon\in(0,\varepsilon_0)$, 有一个 $x(\varepsilon)=x_0+\varepsilon y+r(\varepsilon)\in M$, 并且 $\|r(\varepsilon)\|=o(\varepsilon)$. 于是 $F(x(\varepsilon))=0$. 由于 F 在 x_0 可微, 并且 $F(x_0)=0$, 故

$$0=\varepsilon F'(x_0)y+F'(x_0)r(\varepsilon)+d(\varepsilon),$$

其中 $\|d(\varepsilon)\|=o(\varepsilon)$. 上式两边除以 ε 并让 $\varepsilon\to 0$ 即得 $F'(x_0)y=0$, 从而 $K\subset\{y\in\mathbb{R}^n \mid F'(x_0)y=0\}$.

其次设 $F'(x_0)y=0$. 令 $L=N(F'(x_0))^{\perp}$. 则 $\mathbb{R}^n$ 可作直交和分解

$$\mathbb{R}^n=N(F'(x_0))\oplus L.$$

现在对于上述给定的 $y\in N(F'(x_0))$, 我们求解如下含未知量 $\varepsilon\in\mathbb{R}$ 和 $z\in L$ 的方程:

$$g(\varepsilon,z)\triangleq F(x_0+\varepsilon y+z)=0.$$

注意, 如果记 $F'(x_0)$ 在 L 上的限制为 A, 那么 $A: L\to\mathbb{R}^m$ 是一对一的并且 $AL=\mathbb{R}^m$, 从而逆变换 A^{-1} 存在. 另一方面, 不难看出 $g(0,0)=0$, $g'_z(0,0)=A$. 因此, 根据数学分析中熟知的隐函数定理, $\exists\varepsilon_0>0$, 以及连续可微函数

$r(\cdot):(-\varepsilon_0,\varepsilon_0)\to L$, 使得 $r(0)=0$ 并且 $g(\varepsilon,r(\varepsilon))=0$, 即

$$F(x_0+\varepsilon y+r(\varepsilon))=0,\quad \forall \varepsilon\in(-\varepsilon_0,\varepsilon_0).$$

显然我们有 $\|r(\varepsilon)\|=o(\varepsilon)$. 这正是说 $y\in K$. ∎

定理 5.5.12 设 M 为 $\mathbb{R}^n$ 中的闭凸集, K 为 M 在 x_0 处的切方向锥, 那么

$$K^*=\{x^*\in\mathbb{R}^n \mid \langle x^*,x\rangle\ge\langle x^*,x_0\rangle,\forall x\in M\}.$$

证明 注意对于任意 $x\in M, y\triangleq x-x_0$ 是 M 在 x_0 的切方向. 这是因为由于 M 是凸集, 故对于 $0<\varepsilon<1$, 有

$$x_0+\varepsilon y=(1-\varepsilon)x_0+\varepsilon x\in M.$$

于是若 $x^*\in K^*$, 则

$$\langle x^*,x-x_0\rangle\ge 0,\quad \forall x\in M,$$

即 $\langle x^*,x\rangle\ge\langle x^*,x_0\rangle,\forall x\in M$.

现在设 x^* 满足 $\langle x^*,x-x_0\rangle\ge 0$, $\forall x\in M$. 任取 $y\in K$, 于是 $\exists\varepsilon_0>0$ 使得对于每个 $\varepsilon\in(0,\varepsilon_0)$, 存在 $x(\varepsilon)\in M$, 并且

$$x(\varepsilon)=x_0+\varepsilon y+r(\varepsilon),\quad \|r(\varepsilon)\|=o(\varepsilon).$$

于是

$$\langle x^*,x(\varepsilon)-x_0\rangle\ge 0,\quad \forall 0<\varepsilon<\varepsilon_0.$$

由此我们得到

$$\begin{aligned}\langle x^*,y\rangle&=\frac{\langle x^*,x(\varepsilon)-x_0\rangle}{\varepsilon}-\frac{\langle x^*,r(\varepsilon)\rangle}{\varepsilon}\\&\ge-\frac{\langle x^*,r(\varepsilon)\rangle}{\varepsilon}.\end{aligned}$$

让 $\varepsilon\to 0^+$ 可知 $\langle x^*,y\rangle\ge 0,\forall y\in K$. 因此 $x^*\in K^*$. ∎

注 采用极化锥的记号,定理 5.5.12 是说: $K^o = N_M(x_0)$, 这里 $N_M(x_0)$ 表示 M 在 x_0 处的法向锥. 于是

$$\mathrm{cl}\,K = K^{oo} = N_M(x_0)^o.$$

法向锥的极化锥也叫做切向锥, 记作 $T_M(x_0)$(见习题 4.2). 因此在 M 为闭凸集的情况下, $T_M(x_0)$ 是本节所定义的切方向锥 K 的闭包.

使用上面所得到的结果, 现在我们来导出几种极值问题的必要条件.

(1) 条件极值(等式约束情形):

$$\begin{cases} f_0(x) \to \min, \\ f_i(x) = 0, i = 1, 2, \cdots, m. \end{cases} \tag{5.5.14}$$

定理 5.5.13 设 $f_0, \cdots, f_m$ 是 $\mathbb{R}^n$ 上的可微函数. 设 x_0 是问题 (5.5.14) 的解. 那么存在不全为零的实数 $\lambda_0, \cdots, \lambda_m$, 并且 $\lambda_0 \geq 0$, 使得

$$\lambda_0 f_0'(x_0) + \cdots + \lambda_m f_m'(x_0) = 0. \tag{5.5.15}$$

证明 如果 $f_1'(x_0), \cdots, f_m'(x_0)$ 线性相关, 则存在不全为零的实数 $\mu_1, \cdots, \mu_m$, 使得

$$\mu_1 f_1'(x_0) + \mu_m f_m'(x_0) = 0.$$

于是我们可以取 $\lambda_0 = 0, \lambda_i = \mu_i, i = 1, \cdots, m$, 从而式 (5.5.15) 成立. 因此只要考虑 $f_1'(x_0), \cdots, f_m'(x_0)$ 线性无关的情况. 在这种情况下, 等式约束集合

$$M_1 = \{x \in \mathbb{R}^n \mid f_i(x) = 0, i = 1, \cdots, m\}$$

在 x_0 的的切方向锥 (切子空间) 是

$$K_1 = \{y \in \mathbb{R}^n \mid \langle f_i'(x_0), y\rangle = 0, i = 1, \cdots, m\}.$$

容易验证

$$\begin{aligned} K_1^* &= \operatorname{span}\{f_i'(x_0), i = 1, \cdots, m\} \\ &= \left\{ \sum_{i=1}^m \lambda_i f_i'(x_0) \,\middle|\, \lambda_i \in \mathbb{R}, i = 1, \cdots, m \right\}. \end{aligned}$$

根据定理 5.5.7, 可微函数 f_0 在 x_0 处的减小方向锥是

$$K_0 = \{y \in \mathbb{R}^n \mid \langle f_0'(x_0), y\rangle < 0\}.$$

当 $f_0'(x_0) \neq 0$ 时，其对偶锥是 (见例 3.3.5)

$$K_0^* = \{\lambda_0 f_0'(x_0) \mid \lambda_0 \geq 0\}.$$

(当 $f_0'(x_0) = 0$ 时，我们可以取 $\lambda_0 = 1, \lambda_1 = \cdots = \lambda_m = 0$) 最后利用定理 5.5.4, 即得式 (5.5.15). ∎

注 从定理的证明看到，当 $f_1'(x_0), \cdots, f_m'(x_0)$ 线性无关时， $\lambda_0 > 0$, 从而式 (5.5.15) 可以写成

$$f_0'(x_0) + \lambda_1 f_1'(x_0) + \cdots + \lambda_m f_m'(x_0) = 0.$$

特别在无约束情形下，我们得到通常的无约束极值条件 $f_0'(x_0) = 0$. 数 $\lambda_1, \cdots, \lambda_m$ 通称为 Lagrange 乘子. 这就是数学分析中熟知的 Lagrange 乘子法则.

(2) 规划问题 (不等式约束情形):

$$\begin{cases} f_0(x) \to \min, \\ f_i(x) \leq 0, \quad i = 1, \cdots, m. \end{cases} \tag{5.5.16}$$

定理 5.5.14 设 $f_0, \cdots, f_m$ 是 $\mathbb{R}^n$ 上的函数， x_0 是问题 (5.5.16) 的解. 假定 $f_0, \cdots, f_m$ 在 x_0 处可微. 那么存

在不全为零的实数 $\lambda_0,\cdots,\lambda_m$, 使得

$$\begin{cases} \lambda_0 f_0'(x_0)+\cdots+\lambda_m f_m'(x_0)=0, \\ \lambda_i \geq 0, \quad i=0,1,\cdots,m, \\ \lambda_i f_i(x_0)=0, \quad i=1,\cdots,m. \end{cases} \tag{5.5.17}$$

证明 令

$$M_i=\{x\in \mathbb{R}^n \mid f_i(x)\leq 0\}, \quad i=1,\cdots,m,$$

并设 K_i 是 M_i 在 x_0 处的容许方向锥, $i=1,\cdots,m$. 若 $f_i(x_0)<0$, 则 $K_i=\mathbb{R}^n$; 而若 $f_i(x_0)=0$ 且 $f_i'(x_0)\neq 0$, 则 $K_i=\{y\in \mathbb{R}^n \mid \langle f_i'(x_0),y\rangle<0\}$. 最后若 $f_i(x_0)=0$, 而 $f_i'(x_0)=0$, 则只要取 $\lambda_i=1$, $\lambda_j=0,\forall j\neq i$, 就可知式 (5.5.17) 成立. 于是当 $f_i(x_0)<0$ 时, $K_i^*=\{0\}$, 而当 $f_i(x_0)=0$ 且 $f_i'(x_0)\neq 0$ 时, $K_i^*=\{\lambda_i f_i'(x_0) \mid \lambda_i\geq 0\}$. f_0 在 x_0 处的减小方向锥 K_0 是

$$K_0=\{y\in \mathbb{R}^n \mid \langle f_0'(x_0),y\rangle<0\},$$

并且当 $f_0'(x_0)\neq 0$ 时相应的对偶锥 K_0^* 是

$$K_0^*=\{\lambda_0 f_0'(x_0) \mid \lambda_0\geq 0\}.$$

由此可知式 (5.5.17) 成立. ∎

定理 5.5.15 在定理 5.5.14 的假设下, 如果下列两条件之一成立:

(1) $f_1,\cdots,f_m$ 是凸函数, 并且存在 $\tilde{x}\in \mathbb{R}^n$, 使得 $f_i(\tilde{x})<0, i=1,\cdots,m$;

(2) $f_1,\cdots,f_m$ 是仿射的, 即

$$f_i(x)=\langle x,a_i\rangle-\alpha_i, \quad i=1,\cdots,m.$$

那么 $\lambda_0 > 0$, 从而条件 (5.5.17) 等价于

$$\begin{cases} f_0'(x_0) + \lambda_1 f_1'(x_0) + \cdots + \lambda_m f_m'(x_0) = 0, \\ \lambda_i \geq 0,\ \lambda_i f_i(x_0) = 0, \quad i = 1, \cdots, m. \end{cases} \tag{5.5.18}$$

证明 (1) 从定理 5.5.6 的证明可见, 只要 $\bigcap_{i=1}^{m+1} M_i \neq \emptyset$, 就能保证 $\lambda_0 \neq 0$. 但在目前情况下, 显然 $\tilde{x} - x_0 \in \bigcap_{i=1}^{m+1} M_i$.

(2) 我们令 $M_1 = \{x \in \mathbb{R}^n \mid \langle x, a_i \rangle \leq \alpha_i, i = 1, \cdots, m\}$ (当然可能 $\operatorname{int} M_1 = \emptyset$). M 在 x_0 处的切方向锥 K_1 的对偶锥 K_1^* 是

$$K_1^* = \left\{ \sum_{i=1}^{m} \lambda_i a_i \,\middle|\, \lambda_i \geq 0, \lambda_i(\langle x_0, a_i \rangle - \alpha_i) = 0, i = 1, \cdots, m \right\}.$$

(见习题 5.5.1) 由此利用定理 5.5.4 得证定理结论成立. ∎

注意, 利用定理 5.5.15, 我们立即可得线性规划中的一些主要结果.

(3) 一般数学规划问题

$$\begin{cases} f_0(x) \to \min, \\ f_i(x) \leq 0, \quad i = 1, \cdots, m, \\ f_j(x) = 0, \quad j = m+1, \cdots, m+k, \\ x \in M. \end{cases} \tag{5.5.19}$$

定理 5.5.16 设 $f_0, \cdots, f_{m+k}$ 是 $\mathbb{R}^n$ 上的函数. 假定它们在 x_0 的一个邻域内可微, $M \subset \mathbb{R}^n$ 是含内点的凸集. 那么存在不全为零的实数 $\lambda_0, \cdots, \lambda_{m+k}$, 使得

$$\begin{cases} -x^* = -\sum_{i=0}^{m+k} \lambda_i f_i'(x_0) \in N_M(x_0), \\ \lambda_0 \geq 0, \lambda_i f_i(x_0) = 0, i = 1, \cdots, m. \end{cases} \tag{5.5.20}$$

其中 $N_M(x_0)$ 为 M 在 x_0 处的法向锥.

证明 留作练习. ▮

下一定理指出在更严格的要求下，定理 5.5.4 中的条件不仅必要而且也是充分的.

定理 5.5.17 设 $f_0: \mathbb{R}^n \to \mathbb{R}$ 为凸函数，$M_1, \cdots, M_{m+1}$ 为 $\mathbb{R}^n$ 中的凸集. 假定存在点 $\tilde{x}$ 使得 $\tilde{x} \in \operatorname{int} M_i, i = 1, \cdots, m$, $\tilde{x} \in M_{m+1}$. 设 $x_0 \in M = \bigcap_{i=1}^{m+1} M_i$, 并设 K_0 是 f_0 在 x_0 的减小方向锥，$K_1, \cdots, K_m$ 为 $M_1, \cdots, M_m$ 在 x_0 处的容许方向锥，K_{m+1} 为 M_{m+1} 在 x_0 处的切方向锥. 那么 x_0 为 f_0 在 M 上的极小值点当且仅当存在不全为零的向量 $x_i^* \in K_i^*, i = 0, \cdots, m+1$, 使得

$$x_0^* + \cdots + x_{m+1}^* = 0. \tag{5.5.21}$$

证明 *必要性*: 如果 K_0 非空，则必要性从定理 5.5.4 得出；而若 K_0 为空集，则取任意 $x_i^* \in K_i^*, i = 1, \cdots, m+1$, 并令 $x_0^* = -x_1^* - \cdots - x_{m+1}^*$. 由于 $K_0 = \emptyset$, 故 $K_0^* = \mathbb{R}^n$, 因此 $x_0^* \in K_0^*$.

充分性: 设存在 $x_1 \in M$ 使得 $f_0(x_1) < f_0(x_0)$. 对于 $\lambda \in (0,1)$, 令 $x(\lambda) \triangleq \lambda\tilde{x} + (1-\lambda)x_1$. 由于诸 M_i 均为凸集，故 $x(\lambda) \in M_i, i = 1, \cdots, m+1$; 此外，由于 $\tilde{x} \in \operatorname{int} M_i, i = 1, \cdots, m$, 有 $x(\lambda) \in \operatorname{int} M_i, i = 1, \cdots, m$. 既然 f_0 是 $\mathbb{R}^n$ 上的连续函数，可取 $\lambda > 0$ 充分小，使得 $f_0(x(\lambda)) < f_0(x_0)$. 令 $z = x(\lambda) - x_0$, 则

$$f_0(x_0 + \varepsilon z) \le \varepsilon f_0(x(\lambda)) + (1-\varepsilon) f_0(x_0), \quad 0 < \varepsilon < 1.$$

因此

$$f'_{0+}(x_0; z) = \lim_{\varepsilon \downarrow 0} \frac{f_0(x_0 + \varepsilon z) - f_0(x_0)}{\varepsilon}$$

$$\leq \lim_{\varepsilon \downarrow 0} \frac{\varepsilon(f_0(x(\lambda)) - f_0(x_0))}{\varepsilon}$$
$$= f_0(x(\lambda)) - f_0(x_0) < 0.$$

从而依据定理 5.5.7, $z \in K_0$. 此外,

$$\begin{cases} x_0 + \varepsilon z \in \operatorname{int} M_i, i = 1, \cdots, m, \forall\, 0 < \varepsilon < 1, \\ x_0 + \varepsilon z \in M_{m+1}, \forall\, 0 < \varepsilon < 1, \end{cases}$$

因此 $z \in K_i, i = 1, \cdots, m+1$. 这样我们找到一向量 $z \in \bigcap_{i=0}^{m+1} K_i$, 并且 $K_0, \cdots, K_m$ 为开凸锥. 但根据定理 5.5.3, 这是不可能的. 因此假设 $f_0(x_1) < f_0(x_0)$ 不成立, 从而 x_0 是 f_0 在 M 上的极小值点. ▮

利用上述结果, 我们也可以得到凸规划问题中的结论, 这里不再赘述.

习　题　5.5

5.5.1 设 $M = \{x \in \mathbb{R}^n \mid \langle x, a_i \rangle \leq \alpha_i, i = 1, \cdots, m\}$, 其中 $a_i \in \mathbb{R}^n; \alpha_i \in \mathbb{R}, i = 1, \cdots, m$. 试证: M 在 x_0 处的切方向锥 K 的对偶锥 K^* 是

$$K^* = \left\{ \sum_{i=1}^{m} \lambda_i a_i \,\middle|\, \lambda_i \geq 0, \right.$$
$$\left. \lambda_i(\langle x_0, a_i \rangle - \alpha_i) = 0, i = 1, \cdots, m \right\}.$$

5.5.2 设 $A : \mathbb{R}^n \to \mathbb{R}^m$ 为线性变换. 令

$$K = \{(x, Ax) \in \mathbb{R}^n \times \mathbb{R}^m \mid x \in \mathbb{R}^n\}.$$

试证:

$$K^* = \{(-A^* y^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^m \mid y^* \in \mathbb{R}^m\}.$$

5.5.3 设 K_2 是 $\mathbb{R}^m$ 中一凸锥，$K_1 = \{x \in \mathbb{R}^n \mid Ax \in K_2\}$. 假定如下两条件之一成立：

(1) $\exists \tilde{x} \in \mathbb{R}^n$ 使得 $A\tilde{x} \in \operatorname{int} K_2$;

(2) K_2 为 $\mathbb{R}^m$ 的非负象限.

那么 $K_1^* = A^* K_2$.

5.5.4 完成定理 5.5.16 的证明.

5.5.5 设 $a_1, \cdots, a_m$ 为 $\mathbb{R}^n$ 中的 m 个线性无关元，令

$$K = \{x \in \mathbb{R}^n \mid \langle x, a_i \rangle = 0, i = 1, \cdots, m\}.$$

试证：$K^* = \operatorname{span}\{a_1, \cdots, a_m\}$.

5.5.6 设 $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \cdots, m$. 令

$$f(x) = \max_{1 \le i \le m} f_i(x), \quad x \in \mathbb{R}^n.$$

假定 $f_1, f_2, \cdots, f_m$ 在 x_0 处沿 y 方向可微. 试证：f 在 x_0 处沿 y 方向也可微，并且

$$\begin{cases} f'_+(x_0; y) = \max_{i \in I} f'_{i+}(x_0; y), \\ I = \{i \mid 1 \le i \le m, f_i(x_0) = f(x_0)\}. \end{cases}$$

5.5.7 给定 $a_i \in \mathbb{R}^n, i = 1, \cdots, m$. 令

$$\begin{aligned} K = \{x \in \mathbb{R}^n \mid \langle x, a_i \rangle \ge 0, 1 \le i \le k; \\ \langle x, a_j \rangle = 0, k+1 \le j \le m\}. \end{aligned}$$

试证：

$$K^* = \left\{ \sum_{i=1}^m \lambda_i a_i \,\middle|\, \lambda_i \ge 0, 1 \le i \le k; \lambda_j \in \mathbb{R}, k+1 \le j \le m \right\}.$$

附录 A 凸集与 Brouwer 不动点定理

设 $M \subset \mathbb{R}^n$, $f: M \to \mathbb{R}^n$. 我们称 $x \in M$ 为映射 f 的不动点，是指它满足 $f(x) = x$. 显然，对于任意一个映射 $g: M \to \mathbb{R}^n$, 求解方程 $g(x) = 0$ 都可以归结为求某一个新的映射 f 的不动点 x 的存在性. 例如，我们可以取 $f(x) = x + g(x)$.

Brouwer 不动点定理是近代数学中最重要的成就之一. 这个定理是说： $\mathbb{R}^n$ 中有界闭凸集到其自身的每一个连续映射至少有一个不动点. 定理的叙述和结论都很简单，但其证明却并不是初等的. Brouwer 不动点定理的早期的证明多采用组合拓扑、同伦论或微分形式等方法. 直到 50 年代, Nagumo 等给出了拓扑度理论的解析叙述，为不熟悉拓扑方法的分析学者和对数学知之甚少的工程师等实际工作者理解和学习拓扑度理论以及 Brouwer 不动点定理带来了很大的方便. 但是 Brouwer 不动点定理的证明仍然依赖于相当深奥的拓扑度工具. 此后又有不少数学工作者致力于寻找这个定理的初等证明，终获成功. 这里我们介绍 C. A. Rogers 在 1980 年给出的一个简单证明.

一般地，我们称 $\mathbb{R}^n$ 中的两个集合 A 和 B 是拓扑等价的，是指存在 A 到 B 之上的某个双方一对一的连续映射. 我们先证明一个引理.

引理 A.1 设 K 为 $\mathbb{R}^n$ 中的紧凸集, $\dim K = m$, 那么 K 拓扑等价于 $\mathbb{R}^m$ 中的闭单位球 $\overline{B}_m(0,1)$.

证明 不妨假定 $\dim K = n$，否则的话，作一平移使得 $0 \in K$ 并且在 aff K 中进行讨论. 于是设 K 是有非空内部的紧凸集，并且 $0 \in \text{int } K$(否则作平移). 依引理 3.3.7, K 的 Minkowski 函数 $p_K(\cdot)$ 是连续函数， $K \doteq \{x \in \mathbb{R}^n \mid p_K(x) \leq 1\}$，而且 $\text{int } K = \{x \in \mathbb{R}^n \mid p_K(x) < 1\}$. 因此，

$$\text{bd}\, K = \{x \in \mathbb{R}^n \mid p_K(x) = 1\}.$$

我们定义映射 $T : \mathbb{R}^n \to \mathbb{R}^n$ 如下:

$$T(x) = \begin{cases} (p_K(x)/\|x\|)x, & x \in \mathbb{R}^n \backslash \{0\}, \\ 0, & x = 0, \end{cases}$$

从 p_K 的定义看出， $T(K) \subset \overline{B}(0,1)$，这里 $\overline{B}(0,1)$ 表示 $\mathbb{R}^n$ 中的单位球. 任取 $y \in \overline{B}(0,1), y \neq 0$，容易验证 $x = (\|y\|/p_K(y))y \in K$，并且 $T(x) = y$，这表明 $T(K) = \overline{B}(0,1)$. 由此可见，

$$T^{-1}(y) = \begin{cases} (\|y\|/p_K(y))y, & y \in \mathbb{R}^n \backslash \{0\}, \\ 0, & x = 0. \end{cases}$$

此外 T 还是一对一的. 为了完成证明，剩下只须证明 T 的逆 T^{-1} 也是连续的. 事实上，不难看出存在某个常数 $\beta > 0$ 使得

$$\|y\| \leq \beta p_K(y), \quad \forall\, y \in \mathbb{R}^n. \tag{A.1}$$

从 (A.1) 推出

$$\|T^{-1}(y)\| \leq \beta \|y\|, \quad \forall\, y \mathbb{R}^n.$$

这表明 T^{-1} 在 0 点连续. T^{-1} 在 $y \neq 0$ 处的连续性是 p_K 的连续性的直接结果. ∎

这样，为了证明 Brouwer 不动点定理对于有界闭凸集成立，只要证明对于闭单位球 $\overline{B}(0,1)$ 成立就够了．事实上，设 $f: K \to K$ 为连续映射，并设 $T: \overline{B}(0,1) \to K$ 为双方一对一的连续映射．我们定义新的连续映射 $g: \overline{B}(0,1) \to \overline{B}(0,1)$:

$$g(x) = T^{-1}(f(Tx)), \quad \forall x \in \overline{B}(0,1).$$

于是当 Brouwer 不动点定理对于 $\overline{B}(0,1)$ 成立时，存在 g 的不动点 $x_0 \in \overline{B}(0,1)$ 使得 $g(x_0) = x_0$. 令 $y_0 = T(x_0)$, 则 $y_0 \in K$, 并且 $y_0 = T(g(x_0)) = f(y_0)$, 即 y_0 是 f 在 K 上的不动点．

引理 A.2 设 $f: \overline{B}(0,1) \to S$ 为连续映射，其中 S 为 $\mathbb{R}^n$ 的以 0 为中心的单位球面．那么不可能成立 $f(x) = x, \forall x \in S$, 即 f 限制在单位球面 S 上不可能是恒等映射．

证明 我们分两步来完成证明．

(1) 首先证明，如果 $f(x) = x, \forall x \in S$ 成立，则必有连续可微映射 $h: \overline{B}(0,1) \to S$ 满足 $h(x) = x, \forall x \in S$.

事实上，$f(x) - x$ 在 $\overline{B}(0,1)$ 上连续，并且 $f(x) - x = 0, \forall x \in S$; 又有 $\|f(x) - x\| \le 2, \forall x \in \overline{B}(0,1)$. 因此可以找到一个数 θ, $3/4 < \theta < 1$, 使得

$$\|f(x) - x\| < 1/4, \quad \forall \theta \le \|x\| \le 1.$$

设 $e_1, \cdots, e_n$ 是 $\mathbb{R}^n$ 中沿各标准坐标轴的单位向量．根据 Weierstrass 逼近定理，存在多项式 $p_i(x_1, \cdots, x_n), 1 \le i \le n$, 使得

$$\left\| \sum_{i=1}^{n} p_1(x_1, \cdots, x_n) e_i - (f(x) - x) \right\| < 1/4, \quad \forall x \in \overline{B}(0,1).$$

令 $p(x)=\sum_{i=1}^n p_i(x)e_i$. 再次应用 Weierstrass 逼近定理于单变量函数，总能找到一个多项式 q, 使得

$$\begin{cases} 3/4 \le q(r^2) \le 1, & 0 \le r \le \theta; \\ |q(r^2)| \le 1, & \theta \le r \le 1; \\ q(1)=1. \end{cases}$$

设 $g(x)=x+q(\|x\|^2)p(x)$, 则当 $0\le\|x\|\le\theta$ 时，有

$$\begin{aligned} \|g(x)\| &= \|x+q(\|x\|^2)p(x)\| \\ &= \|f(x)+q(\|x\|^2)\{p(x)-f(x)+x\} \\ &\quad +\{q(\|x\|^2)-1\}\{f(x)-x\}\| \\ &\ge \|f(x)\|-|q(\|x\|^2)|\|p(x)-f(x)+x\| \\ &\quad -|1-q(\|x\|^2)|\|f(x)-x\| \\ &\ge 1-1/4-2/4=1/4. \end{aligned}$$

类似地，当 $\theta\le\|x\|\le 1$ 时，有

$$\begin{aligned} \|g(x)\| &= \|x+q(\|x\|^2)\{p(x)-f(x)+x\} \\ &\quad +q(\|x\|^2)\{f(x)-x\}\| \\ &\ge \|x\|-|q(\|x\|^2)|\{\|p(x)-f(x)+x\|+\|f(x)-x\|\} \\ &\ge \theta-(1/4+1/4)\ge 1/4. \end{aligned}$$

因此

$$\begin{cases} \|g(x)\|\ge 1/4, & \forall\, x\in\overline{B}(0,1), \\ g(x)=x, & \forall\, x\in S. \end{cases}$$

注意 $g(x)$ 的每个分量是 $x_1,\cdots,x_n$ 的多项式，故 g 是连续可微的. 从而若令

$$h(x)=g(x)/\|g(x)\|, \quad \forall\, x\in\overline{B}(0,1),$$

则 $h:\overline{B}(0,1)\to S$ 是连续可微的映射，并且 $h(x)=x,\forall x\in S$.

(2) 现在我们从上面得到的函数 $h(x)$ 来推出矛盾. 为此令

$$h_t(x)=x+t(h(x)-x),\ \ 0\le t\le 1,\ x\in\overline{B}(0,1).$$

由于 h 连续可微，故 $\varphi(x)=h(x)-x$ 也连续可微，并且存在常数 $c>0$, 使得

$$\|\varphi(x)-\varphi(y)\|\le c\|x-y\|,\quad ,\forall\, x,y\in\overline{B}(0,1).$$

注意，当 $0\le t<1/c$ 时， $h_t:\overline{B}(0,1)\to\overline{B}(0,1)$ 是一对一的，因为如果 $h_t(x)=h_t(y)$, 并且 $0\le t<1/c$, 则

$$\|x-y\|=t\|\varphi(x)-\varphi(y)\|\le tc\|x-y\|.$$

从而 $x=y$.

又注意 $\varphi(x)$ 关于 $x_1,\cdots,x_n$ 的偏微商在 $\overline{B}(0,1)$ 上是一致有界的，因此存在一充分小的正数 t_0, 使得当 $0\le t\le t_0$ 时， $h_t(x)$ 的 Jacobi 矩阵

$$h_t'(x)=I_n+t\varphi'(x),\quad \forall\, x\in\overline{B}(0,1)\qquad\qquad (A.2)$$

是非奇异的，这里 I_n 是 $n\times n$ 单位矩阵. 于是对于 $t\in[0,t_0]$, 根据隐函数定理， h_t 把 $B(0,1)$ 映成 $B(0,1)$ 中的一个开连通集 G_t. 对于任意固定的 $t\in[0,t_0]$, 考察点 $z\in\overline{B}(0,1)\backslash G_t$. 任取点 $w\in G_t$, 设 b 为 z 和 w 的连线与 G_t 的边界的一个交点. 由于 $h_t(\overline{B}(0,1))$ 是紧集，必有 $x\in\overline{B}(0,1)$ 使得 $b=h_t(x)$. 但 $b\notin G_t$, 故 x 不在 $\overline{B}(0,1)$ 的内部，即 $\|x\|=1$. 于是 $b=x$, 并且 z 和 b 位于 $\overline{B}(0,1)$ 的边界. 由于 $h_t:S\to S$, 可见当 $0\le t\le t_0$ 时， h_t 是 $\overline{B}(0,1)$ 到其自身之上的双方一对一的映射.

令对于 $0 \le t \le 1$, 定义积分

$$J(t) = \int_{\overline{B}(0,1)} \det h_t'(x)\, dx.$$

当 $0 \le t \le t_0$ 时, $J(t)$ 恰好是单位球 $\overline{B}(0,1)$ 的体积 V_n. 但从 (A.2) 显然可见 $J(t)$ 是 t 的多项式. 因此 $J(t)$ 只能是常值, 即 $J(t) = V_n$, $0 \le t \le 1$. 但是,

$$\langle h_1(x), h_1(x)\rangle = 1, \quad \forall x \in \overline{B}(0,1),$$

所以

$$\left\langle \frac{\partial h_1(x)}{\partial x_i}, h_1(x) \right\rangle = 0, \quad \forall x \in \overline{B}(0,1), 1 \le i \le n.$$

由此可知

$$\det h_t'(x) = 0, \quad \forall x \in \overline{B}(0,1),$$

从而 $J(1) = 0$, 与上述结论 $J(1) = V_n$ 相矛盾. ∎

定理 A.3 (Brouwer定理) 设 $f : \overline{B}(0,1) \to \overline{B}(0,1)$ 为连续映射. 那么 f 在 $\overline{B}(0,1)$ 中至少有一个不动点.

证明 用反证法. 设 f 在 $\overline{B}(0,1)$ 中没有不动点. 于是 $g(x) = x - f(x) \ne 0, \forall x \in S$, 并且 $g(x)$ 在 S 上指向外, 即

$$\langle x, g(x)\rangle = 1 - \langle x, f(x)\rangle > 0, \quad \forall x \in S. \tag{A.3}$$

令

$$\varphi(x) = x - \lambda(x) f(x), \quad , x \in \overline{B}(0,1),$$

其中

$$\lambda(x) = \frac{1 - \langle x, x\rangle}{1 - \langle x, f(x)\rangle}, \quad x \in \overline{B}(0,1). \tag{A.4}$$

那么 $\varphi(x)$ 在 $\overline{B}(0,1)$ 中连续, 并且在 S 上 $\varphi(x) = x$. 今证 $\varphi(x) \ne 0, \forall x \in \overline{B}(0,1)$. 事实上, 若有 $x \in \overline{B}(0,1)$ 使得

$\varphi(x) = 0$, 则 $x = \lambda f(x)$. 但据此从 (A.4) 可知 $\lambda = 1$, 即 $x = f(x)$, 这与 f 在 $\overline{B}(0,1)$ 上无不动点相矛盾. 于是我们可以找到连续映射 $\psi(x) = \varphi(x)/\|\varphi(x)\|$, 使得

$$\begin{cases} \psi : \overline{B}(0,1) \to S; \\ \psi(x) = x, \quad \forall x \in S. \end{cases}$$

但从引理 A.2 可知这是不可能的. 从而 f 在 $\overline{B}(0,1)$ 上必有不动点. ∎

参考文献

[1] 史树中，凸分析，上海科学技术出版社， 1990.

[2] 周鸿兴，应用实分析基础，科学出版社， 1991.

[3] 关肇直、韩京清、秦化淑、王朝珠、王世林，极值控制与极大值原理，科学出版社， 1980.

[4] 张建中、许绍吉，线性规划，科学出版社， 1990.

[5] Barbu, V. & Th. Precupanu, Convexity and Optimization in Banach Spaces, Romania International Publisher, Bucuresti, 1978.

[6] Danski, J. N., The Theory of Maxin, Springer, Berlin, 1967.

[7] Girsanov, I. V. ; Lectures on Mathematical Theory of Extremum Problems, Springer–Verlag, Berlin–Heidelberg–New York, 1972.

[8] Golstein, E. G. , Theory of Convex Programming, American Math. Soc., Providence, Rhode Island, 1972.

[9] Grünbaum, B. , Convex Polytopes, Wiley, New York, 1967.

[10] McMullen, M. & G. C. Shepard, Convex Polytopes and the Upper Bound Conjecture, Cambridge University Press, Cambridge, England, 1971.

[11] Lay, S. R. , Convex Sets and Their Applications, John Wiley & Sons, New York–Chichester–Brisbane–Toronto–Singapore, John Wiley & Sons, 1982.

[12] Rockafellar, R. T. , Convex Analysis, Princeton University Press, Princeton, New Jersey, 1970.

[13] Stoer, J. & Ch. Witzgall, Convexity and Optimization in Finite Dimensions I, Spinger–Verlag, Berlin–Heidelberg–New York, 1970.

名 词 索 引

其他